

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 543.069 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-11
Perfect score: 20
Sequence: 1 cgaggcgccagcccccacaa 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residus

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|------------|---------------------|
| 1 | 19 | 95.0 | 17131 | 1 | AE007080 | AE007080 Mycobacte |
| 2 | 19 | 95.0 | 306550 | 1 | EX248342 | EX248342 Mycobacte |
| 3 | 19 | 95.0 | 348247 | 15 | EX842579 | EX842579 Mycobacte |
| C 4 | 17.4 | 87.0 | 1073 | 9 | RMU22370 | U22370 Rhizobium m |
| C 5 | 17.4 | 87.0 | 2253 | 9 | AK098077 | AK098077 Homo sapi |
| C 6 | 17.4 | 87.0 | 10029 | 1 | AE007196 | AE007196 Sinorhizo |
| C 7 | 17.4 | 87.0 | 10296 | 1 | AE007297 | AE007297 Sinorhizo |
| C 8 | 17.4 | 87.0 | 10578 | 1 | AE007314 | AE007314 Sinorhizo |
| 9 | 17.4 | 87.0 | 10931 | 1 | AE007285 | AE007285 Sinorhizo |
| 10 | 17.4 | 87.0 | 12706 | 4 | BTINXGENE | Y11915 B.taurus mR |
| 11 | 17.4 | 87.0 | 37476 | 9 | HS316G12 | AL031709 Human DNA |
| C 12 | 17.4 | 87.0 | 86834 | 9 | AF003626 | AF003626 Homo sapi |
| C 13 | 17.4 | 87.0 | 97834 | 2 | AL354677 | AL354677 Homo sapi |
| 14 | 17.4 | 87.0 | 105589 | 2 | AC092269 | AC092269 Homo sapi |
| C 15 | 17.4 | 87.0 | 129679 | 9 | AF187000 | AF187000 Homo sapi |
| C 16 | 17.4 | 87.0 | 154807 | 2 | AC069493 | AC069493 Homo sapi |
| C 17 | 17.4 | 87.0 | 166847 | 9 | AC008393 | AC008393 Homo sapi |
| 18 | 17.4 | 87.0 | 175511 | 2 | AC023924 | AC023924 Homo sapi |
| 19 | 17.4 | 87.0 | 181070 | 2 | AC142187 | AC142187 Rattus no |
| 20 | 17.4 | 87.0 | 182765 | 2 | AC124790 | AC124790 Homo sapi |
| 21 | 17.4 | 87.0 | 186908 | 2 | AC124791 | AC124791 Homo sapi |
| 22 | 17.4 | 87.0 | 189333 | 1 | RME603647 | AL603647 Rhizobium |
| C 23 | 17.4 | 87.0 | 189436 | 9 | AF274573 | AF274573 Homo sapi |
| C 24 | 17.4 | 87.0 | 195563 | 9 | AC100872 | AC100872 Homo sapi |
| C 25 | 17.4 | 87.0 | 197000 | 2 | AC124859 | AC124859 Homo sapi |
| C 26 | 17.4 | 87.0 | 202437 | 2 | AC023224 | AC023224 Homo sapi |
| 27 | 17.4 | 87.0 | 260050 | 1 | RME603643 | AL603643 Rhizobium |
| 28 | 17.4 | 87.0 | 279011 | 9 | AE006467 | AE006467 Homo sapi |
| 29 | 17.4 | 87.0 | 286550 | 1 | SME591785 | AL591785 Sinorhizo |
| C 30 | 17.4 | 87.0 | 286550 | 1 | SME591785 | AL591785 Sinorhizo |
| C 31 | 17.4 | 87.0 | 299350 | 1 | SME591786 | AL591786 Sinorhizo |
| 32 | 17.4 | 87.0 | 300000 | 1 | SME591784 | AL591784 Sinorhizo |
| 33 | 17.4 | 87.0 | 333800 | 1 | SME591792 | AL591792 Sinorhizo |
| 34 | 17 | 85.0 | 1867 | 9 | AF013956 | AF013956 Homo sapi |
| C 35 | 17 | 85.0 | 88655 | 9 | AC079135 | AC079135 Homo sapi |
| 36 | 17 | 85.0 | 118404 | 4 | AL773591 | AL773591 Pig DNA s |
| C 37 | 17 | 85.0 | 175127 | 2 | AC016048 | AC016048 Homo sapi |
| C 38 | 17 | 85.0 | 189050 | 2 | AC142566 | AC142566 Bos tauru |
| C 39 | 17 | 85.0 | 190050 | 1 | AL646059 | AL646059 Ralstonia |
| C 40 | 17 | 85.0 | 199827 | 2 | AC026373 | AC026373 Homo sapi |
| 41 | 17 | 85.0 | 206960 | 2 | AC128121 | AC128121 Rattus no |
| 42 | 17 | 85.0 | 209751 | 9 | AC087645 | AC087645 Homo sapi |
| C 43 | 17 | 85.0 | 214485 | 2 | AC123416 | AC123416 Rattus no |
| C 44 | 17 | 85.0 | 215313 | 9 | AC100791 | AC100791 Homo sapi |
| C 45 | 17 | 85.0 | 229426 | 2 | AC010532 | AC010532 Homo sapi |
| C 46 | 17 | 85.0 | 246660 | 2 | AC115225 | AC115225 Rattus no |
| C 47 | 17 | 85.0 | 268510 | 2 | AC095840 | AC095840 Rattus no |
| 48 | 17 | 85.0 | 320746 | 2 | AC128373 | AC128373 Rattus no |
| C 49 | 16.8 | 84.0 | 442 | 9 | HS4236689 | AX523503 Sequence |
| C 50 | 16.8 | 84.0 | 993 | 6 | AX497801 | AX497801 Sequence |
| C 51 | 16.8 | 84.0 | 1002 | 6 | AF380137 | AF380137 Takifugu |
| 52 | 16.8 | 84.0 | 4588 | 5 | PD03453 | U34353 Paracoccus |
| 53 | 16.8 | 84.0 | 8851 | 1 | AX647145 | AX647145 Sequence |
| 54 | 16.8 | 84.0 | 9851 | 6 | AE001247 | AE001247 Treponema |
| 55 | 16.8 | 84.0 | 13860 | 1 | AY081135 | AY081135 Takifugu |
| 56 | 16.8 | 84.0 | 22203 | 5 | AC019899 | AC019899 Drosophila |
| 57 | 16.8 | 84.0 | 25155 | 2 | AC019899 | U41022 Caenorhabdi |
| 58 | 16.8 | 84.0 | 35728 | 3 | U41022 | AX708626 Sequence |
| C 59 | 16.8 | 84.0 | 37113 | 6 | AC126346 | AC126346 Homo sapi |
| 60 | 16.8 | 84.0 | 63964 | 2 | HS754E20A | AL096761 Human DNA |
| C 61 | 16.8 | 84.0 | 65201 | 9 | HS436C18 | AL031301 Homo sapi |
| 62 | 16.8 | 84.0 | 122732 | 2 | HSAC001236 | AC001236 Homo sapi |
| 63 | 16.8 | 84.0 | 138229 | 2 | AC002536 | AC002536 Human Chr |
| C 64 | 16.8 | 84.0 | 140977 | 9 | AC129929 | AC129929 Homo sapi |
| 65 | 16.8 | 84.0 | 141449 | 9 | AC129929 | |

| | | | | | | |
|-------|------|------|--------|----|------------|----------------------|
| C 66 | 16.8 | 84.0 | 143565 | 2 | AC023195 | AC023195 Homo sapi |
| C 67 | 16.8 | 84.0 | 144461 | 2 | AC145808 | AC145808 Oryza sat |
| C 68 | 16.8 | 84.0 | 144902 | 2 | AC016114 | AC016114 Homo sapi |
| C 69 | 16.8 | 84.0 | 145042 | 9 | AC091012 | AC091012 Homo sapi |
| C 70 | 16.8 | 84.0 | 147505 | 9 | CNS01012 | AL132708 Human chr |
| C 71 | 16.8 | 84.0 | 148083 | 9 | AF111170 | AF111170 Homo sapi |
| C 72 | 16.8 | 84.0 | 153250 | 3 | AC008332 | AC008332 Drosophila |
| C 73 | 16.8 | 84.0 | 157566 | 9 | AC022317 | AC022317 Homo sapi |
| C 74 | 16.8 | 84.0 | 158467 | 9 | CNS07B5 | AL122556 Human chr |
| C 75 | 16.8 | 84.0 | 162144 | 9 | CNS06CB5 | AL391261 Human chr |
| C 76 | 16.8 | 84.0 | 166380 | 9 | AC087237 | AC087237 Homo sapi |
| C 77 | 16.8 | 84.0 | 174811 | 2 | AC013259 | AC013259 Homo sapi |
| C 78 | 16.8 | 84.0 | 181004 | 9 | CNS01D7 | AL132990 Human chr |
| C 79 | 16.8 | 84.0 | 181683 | 2 | AC021618 | AC021618 Homo sapi |
| C 80 | 16.8 | 84.0 | 184556 | 9 | CNS05TDB | AL356021 Human chr |
| C 81 | 16.8 | 84.0 | 190301 | 3 | AC009352 | AC009352 Drosophila |
| C 82 | 16.8 | 84.0 | 197426 | 3 | CNS07EG9 | AL590442 Chromosome |
| C 83 | 16.8 | 84.0 | 198273 | 9 | CNS07EP4 | AL512355 Human chr |
| C 84 | 16.8 | 84.0 | 254309 | 3 | AE003638 | AE003638 Drosophila |
| C 85 | 16.8 | 84.0 | 346362 | 1 | BE640439 | BE640439 Bordetella |
| C 86 | 16.8 | 84.0 | 348257 | 1 | BE640425 | BE640425 Bordetella |
| C 87 | 16.8 | 84.0 | 349726 | 1 | BE640421 | BE640421 Bordetella |
| C 88 | 16.4 | 82.0 | 354 | 6 | AE988401 | AE988401 Sequence |
| C 89 | 16.4 | 82.0 | 354 | 6 | BD033934 | BD033934 Sequence |
| C 90 | 16.4 | 82.0 | 674 | 5 | AF095342 | AF095342 Oreophryn |
| C 91 | 16.4 | 82.0 | 2230 | 9 | BC013116 | BC013116 Homo sapi |
| C 92 | 16.4 | 82.0 | 2750 | 8 | AK110063 | AK110063 Oryza sat |
| C 93 | 16.4 | 82.0 | 3145 | 6 | BD275095 | BD275095 47 Human |
| C 94 | 16.4 | 82.0 | 3287 | 6 | AX301205 | AX301205 Sequence |
| C 95 | 16.4 | 82.0 | 3349 | 6 | AX405762 | AX405762 Sequence |
| C 96 | 16.4 | 82.0 | 3349 | 6 | HS4427922 | HS4427922 Homo sapi |
| C 97 | 16.4 | 82.0 | 3370 | 6 | AE833184 | AE833184 Sequence |
| C 98 | 16.4 | 82.0 | 3370 | 6 | AK094738 | AK094738 Homo sapi |
| C 99 | 16.4 | 82.0 | 11219 | 1 | AE005981 | AE005981 Caulobact |
| C 100 | 16.4 | 82.0 | 20000 | 9 | AB006684 | AB006684 Homo sapi |
| C 101 | 16.4 | 82.0 | 33110 | 9 | AL354671 | AL354671 Human DNA |
| C 102 | 16.4 | 82.0 | 36284 | 9 | HS495610 | HS495610 Homo sapi |
| C 103 | 16.4 | 82.0 | 39374 | 9 | AC005256 | AC005256 Homo sapi |
| C 104 | 16.4 | 82.0 | 39531 | 9 | AF024534 | AF024534 Homo sapi |
| C 105 | 16.4 | 82.0 | 39782 | 2 | AC138635 | AC138635 Homo sapi |
| C 106 | 16.4 | 82.0 | 40476 | 2 | AC140538 | AC140538 Homo sapi |
| C 107 | 16.4 | 82.0 | 40476 | 2 | AC002116 | AC002116 Human DNA |
| C 108 | 16.4 | 82.0 | 42133 | 9 | AP001060 | AP001060 Homo sapi |
| C 109 | 16.4 | 82.0 | 43444 | 9 | CH19R31491 | CH19R31491 Homo sapi |
| C 110 | 16.4 | 82.0 | 44228 | 2 | AC140533 | AC140533 Homo sapi |
| C 111 | 16.4 | 82.0 | 44635 | 2 | AC140539 | AC140539 Homo sapi |
| C 112 | 16.4 | 82.0 | 45185 | 2 | AC011166 | AC011166 Homo sapi |
| C 113 | 16.4 | 82.0 | 47877 | 9 | HS206C7 | HS206C7 Human DNA |
| C 114 | 16.4 | 82.0 | 48894 | 2 | AC080102 | AC080102 Homo sapi |
| C 115 | 16.4 | 82.0 | 6802 | 2 | AC118753 | AC118753 Homo sapi |
| C 116 | 16.4 | 82.0 | 76142 | 2 | AC139539 | AC139539 Homo sapi |
| C 117 | 16.4 | 82.0 | 8896 | 9 | AC090830 | AC090830 Homo sapi |
| C 118 | 16.4 | 82.0 | 88780 | 2 | AC138711 | AC138711 Homo sapi |
| C 119 | 16.4 | 82.0 | 100976 | 9 | HS1005F21 | HS1005F21 Human DNA |
| C 120 | 16.4 | 82.0 | 107467 | 9 | HS958B3 | HS958B3 Human DNA |
| C 121 | 16.4 | 82.0 | 110000 | 2 | AC003656_5 | Continuation of 6 of |
| C 122 | 16.4 | 82.0 | 119931 | 9 | AC110285 | AC110285 Homo sapi |
| C 123 | 16.4 | 82.0 | 129586 | 9 | AC005086 | AC005086 Homo sapi |
| C 124 | 16.4 | 82.0 | 143192 | 9 | AL954247 | AL954247 Pan trogl |
| C 125 | 16.4 | 82.0 | 145256 | 9 | AL360182 | AL360182 Human DNA |
| C 126 | 16.4 | 82.0 | 149567 | 9 | AL590226 | AL590226 Human DNA |
| C 127 | 16.4 | 82.0 | 158365 | 9 | AC091171 | AC091171 Homo sapi |
| C 128 | 16.4 | 82.0 | 159064 | 9 | AC138512 | AC138512 Homo sapi |
| C 129 | 16.4 | 82.0 | 166878 | 2 | AC023288 | AC023288 Homo sapi |
| C 130 | 16.4 | 82.0 | 167809 | 2 | AC126415 | AC126415 Mus muscu |
| C 131 | 16.4 | 82.0 | 170959 | 2 | AC022197 | AC022197 Homo sapi |
| C 132 | 16.4 | 82.0 | 174235 | 9 | AC145919 | AC145919 Pan trogl |
| C 133 | 16.4 | 82.0 | 175691 | 2 | AC007615 | AC007615 Homo sapi |
| C 134 | 16.4 | 82.0 | 176502 | 2 | AC134706 | AC134706 Rattus no |
| C 135 | 16.4 | 82.0 | 177432 | 9 | AC118274 | AC118274 Homo sapi |
| C 136 | 16.4 | 82.0 | 181714 | 2 | AC090442 | AC090442 Homo sapi |
| C 137 | 16.4 | 82.0 | 184050 | 2 | AC087819 | AC087819 Homo sapi |
| C 138 | 16.4 | 82.0 | 184444 | 2 | AC090471 | AC090471 Homo sapi |
| C 139 | 16.4 | 82.0 | 187824 | 9 | AP001095 | AP001095 Homo sapi |
| C 140 | 16.4 | 82.0 | 189628 | 2 | AC146318 | AC146318 Gallus ga |
| C 141 | 16.4 | 82.0 | 191456 | 2 | AC009130 | AC009130 Homo sapi |
| C 142 | 16.4 | 82.0 | 191541 | 4 | AC144339 | AC144339 Oryctolag |
| C 143 | 16.4 | 82.0 | 194020 | 9 | AC096679 | AC096679 Pan trogl |
| C 144 | 16.4 | 82.0 | 195791 | 2 | AC019281 | AC019281 Homo sapi |
| C 145 | 16.4 | 82.0 | 197605 | 9 | AC096770 | AC096770 Homo sapi |
| C 146 | 16.4 | 82.0 | 202453 | 2 | AC069094 | AC069094 Homo sapi |
| C 147 | 16.4 | 82.0 | 203765 | 2 | AC139635 | AC139635 Gallus ga |
| C 148 | 16.4 | 82.0 | 204433 | 9 | AC09489 | AC09489 Homo sapi |
| C 149 | 16.4 | 82.0 | 207997 | 9 | AC009093 | AC009093 Homo sapi |
| C 150 | 16.4 | 82.0 | 216498 | 2 | AC123722 | AC123722 Mus muscu |
| C 151 | 16.4 | 82.0 | 225204 | 2 | AC099219 | AC099219 Rattus no |
| C 152 | 16.4 | 82.0 | 226460 | 2 | AC114157 | AC114157 Rattus no |
| C 153 | 16.4 | 82.0 | 232450 | 2 | AC095706 | AC095706 Rattus no |
| C 154 | 16.4 | 82.0 | 233479 | 2 | AC126758 | AC126758 Homo sapi |
| C 155 | 16.4 | 82.0 | 233772 | 2 | AC115403 | AC115403 Rattus no |
| C 156 | 16.4 | 82.0 | 236608 | 10 | AC115763 | AC115763 Mus muscu |
| C 157 | 16.4 | 82.0 | 256607 | 2 | AC021040 | AC021040 Homo sapi |
| C 158 | 16.4 | 82.0 | 269607 | 2 | AC145355 | AC145355 Gorilla g |
| C 159 | 16.4 | 82.0 | 340000 | 9 | AP001753 | AP001753 Homo sapi |
| C 160 | 16.4 | 82.0 | 340000 | 9 | AP001754 | AP001754 Homo sapi |
| C 161 | 16.4 | 82.0 | 3417 | 6 | AX481394 | AX481394 Sequence |
| C 162 | 16.4 | 82.0 | 3845 | 6 | AL6753 | AL6753 tie recepto |
| C 163 | 16.4 | 82.0 | 3845 | 6 | AR074868 | AR074868 Sequence |
| C 164 | 16.4 | 82.0 | 3845 | 6 | AX587640 | AX587640 Sequence |
| C 165 | 16.4 | 82.0 | 3845 | 6 | AX671046 | AX671046 Sequence |
| C 166 | 16.4 | 82.0 | 3845 | 9 | HSTIEMR | HSTIEMR Human tie m |
| C 167 | 16.4 | 82.0 | 3960 | 9 | BC038239 | BC038239 Homo sapi |
| C 168 | 16.4 | 82.0 | 6074 | 9 | AF067971 | AF067971 Macaca mu |
| C 169 | 16.4 | 82.0 | 8670 | 6 | BD129605 | BD129605 Polynucle |
| C 170 | 16.4 | 82.0 | 8670 | 8 | CRARG7 | CRARG7 Chlamydomon |
| C 171 | 16.4 | 82.0 | 35716 | 10 | AL928570 | AL928570 Mouse DNA |
| C 172 | 16.4 | 82.0 | 100340 | 2 | AL18286 | AL18286 Oryza sat |
| C 173 | 16.4 | 82.0 | 102286 | 9 | AC087644 | AC087644 Homo sapi |
| C 174 | 16.4 | 82.0 | 106168 | 9 | AC011450 | AC011450 Homo sapi |
| C 175 | 16.4 | 82.0 | 129336 | 9 | AC130457 | AC130457 Homo sapi |
| C 176 | 16.4 | 82.0 | 135417 | 9 | HSDJ41217 | HSDJ41217 Human DNA |
| C 177 | 16.4 | 82.0 | 135996 | 2 | AL356007 | AL356007 Homo sapi |
| C 178 | 16.4 | 82.0 | 140366 | 9 | AC104425 | AC104425 Homo sapi |
| C 179 | 16.4 | 82.0 | 144988 | 10 | AC101022 | AC101022 Mus muscu |
| C 180 | 16.4 | 82.0 | 146514 | 2 | AC141983 | AC141983 Rattus no |
| C 181 | 16.4 | 82.0 | 149480 | 6 | AX329775 | AX329775 Sequence |
| C 182 | 16.4 | 82.0 | 149480 | 6 | AX329776 | AX329776 Sequence |
| C 183 | 16.4 | 82.0 | 149480 | 6 | AX336193 | AX336193 Sequence |
| C 184 | 16.4 | 82.0 | 149480 | 9 | HUU95740 | HUU95740 Human Chrom |
| C 185 | 16.4 | 82.0 | 149480 | 9 | AC130455 | AC130455 Homo sapi |
| C 186 | 16.4 | 82.0 | 150129 | 2 | AC023816 | AC023816 Homo sapi |
| C 187 | 16.4 | 82.0 | 153857 | 2 | AF336380 | AF336380 Mus muscu |
| C 188 | 16.4 | 82.0 | 168813 | 9 | AC013717 | AC013717 Homo sapi |
| C 189 | 16.4 | 82.0 | 170106 | 9 | AC104371 | AC104371 Homo sapi |
| C 190 | 16.4 | 82.0 | 171140 | 9 | AC090685 | AC090685 Homo sapi |
| C 191 | 16.4 | 82.0 | 171140 | 9 | AC090685 | AC090685 Homo sapi |
| C 192 | 16.4 | 82.0 | 172622 | 2 | AC024065 | AC024065 Homo sapi |
| C 193 | 16.4 | 82.0 | 175611 | 2 | AC020621 | AC020621 Homo sapi |
| C 194 | 16.4 | 82.0 | 178733 | 2 | AC068786 | AC068786 Homo sapi |
| C 195 | 16.4 | 82.0 | 181603 | 3 | AC093189 | AC093189 Pan trogl |
| C 196 | 16.4 | 82.0 | 183067 | 9 | AC040914 | AC040914 Homo sapi |
| C 197 | 16.4 | 82.0 | 186643 | 2 | AC144995 | AC144995 Rattus no |
| C 198 | 16.4 | 82.0 | 190183 | 9 | AC100768 | AC100768 Homo sapi |
| C 199 | 16.4 | 82.0 | 190948 | 2 | AC093138 | AC093138 Pan trogl |
| C 200 | 16.4 | 82.0 | 191141 | 2 | AC146230 | AC146230 Homo sapi |
| C 201 | 16.4 | 82.0 | 191530 | 2 | AC140519 | AC140519 Homo sapi |
| C 202 | 16.4 | 82.0 | 193766 | 9 | AC093420 | AC093420 Homo sapi |
| C 203 | 16.4 | 82.0 | 194981 | 9 | AC145997 | AC145997 Pan trogl |
| C 204 | 16.4 | 82.0 | 210688 | 2 | AC139257 | AC139257 Homo sapi |
| C 205 | 16.4 | 82.0 | 212387 | 2 | AC13473 | AC13473 Mus muscu |
| C 206 | 16.4 | 82.0 | 220641 | 2 | AC147123 | AC147123 Pan trogl |
| C 207 | 16.4 | 82.0 | 221062 | 2 | AC098570 | AC098570 Homo sapi |
| C 208 | 16.4 | 82.0 | 226841 | 2 | HSAC002043 | HSAC002043 Rattus no |
| C 209 | 16.4 | 82.0 | 240453 | 2 | AC119387 | AC119387 Rattus no |
| C 210 | 16.4 | 82.0 | 240453 | 2 | AC124895 | AC124895 Rattus no |
| C 211 | 16.4 | 82.0 | 246240 | 2 | AC124895 | AC124895 Rattus no |

| | | | | | | |
|-----|------|------|--------|----|------------|--------------|
| 212 | 16 | 80.0 | 250608 | 2 | AC110700 | Rattus no |
| 213 | 16 | 80.0 | 346259 | 1 | BX540435 | Bordetell |
| 214 | 15.8 | 79.0 | 454 | 6 | AX577678 | Sequence |
| 215 | 15.8 | 79.0 | 477 | 6 | AR427389 | Sequence |
| 216 | 15.8 | 79.0 | 477 | 6 | BD122942 | EST and e |
| 217 | 15.8 | 79.0 | 504 | 6 | AX577682 | Sequence |
| 218 | 15.8 | 79.0 | 523 | 6 | AX873301 | Sequence |
| 219 | 15.8 | 79.0 | 523 | 6 | BD153963 | Primer fo |
| 220 | 15.8 | 79.0 | 578 | 9 | HSXEV1 | |
| 221 | 15.8 | 79.0 | 598 | 8 | AX064351 | H.sapiens X |
| 222 | 15.8 | 79.0 | 636 | 9 | HUMU73B11 | Erlosyce |
| 223 | 15.8 | 79.0 | 786 | 6 | AF087969 | Homo sapi |
| 224 | 15.8 | 79.0 | 786 | 6 | AX474216 | Sequence |
| 225 | 15.8 | 79.0 | 786 | 6 | AX474218 | Sequence |
| 226 | 15.8 | 79.0 | 786 | 6 | AX474220 | Sequence |
| 227 | 15.8 | 79.0 | 786 | 6 | AX474222 | Sequence |
| 228 | 15.8 | 79.0 | 786 | 6 | AX474234 | Sequence |
| 229 | 15.8 | 79.0 | 786 | 6 | AX474237 | Sequence |
| 230 | 15.8 | 79.0 | 786 | 6 | BD185471 | Peptide m |
| 231 | 15.8 | 79.0 | 786 | 6 | BD185472 | Peptide m |
| 232 | 15.8 | 79.0 | 786 | 6 | BD185473 | Peptide m |
| 233 | 15.8 | 79.0 | 786 | 6 | BD185474 | Peptide m |
| 234 | 15.8 | 79.0 | 786 | 6 | BD185480 | Peptide m |
| 235 | 15.8 | 79.0 | 836 | 3 | CGU40207 | Peptide m |
| 236 | 15.8 | 79.0 | 947 | 6 | AX367930 | Sequence |
| 237 | 15.8 | 79.0 | 947 | 6 | AX069385 | Sequence |
| 238 | 15.8 | 79.0 | 947 | 6 | AX335617 | Sequence |
| 239 | 15.8 | 79.0 | 947 | 6 | AX410772 | Sequence |
| 240 | 15.8 | 79.0 | 947 | 6 | HSU73514 | Human short |
| 241 | 15.8 | 79.0 | 960 | 6 | AX053564 | Sequence |
| 242 | 15.8 | 79.0 | 961 | 9 | BC000372 | Homo sapi |
| 243 | 15.8 | 79.0 | 965 | 9 | BC008708 | Homo sapi |
| 244 | 15.8 | 79.0 | 968 | 9 | AF069134 | Homo sapi |
| 245 | 15.8 | 79.0 | 973 | 9 | AF031555 | Homo sapi |
| 246 | 15.8 | 79.0 | 990 | 9 | HSU96132 | Homo sapien |
| 247 | 15.8 | 79.0 | 1005 | 6 | AX053562 | Sequence |
| 248 | 15.8 | 79.0 | 1054 | 6 | BD082375 | Sequence |
| 249 | 15.8 | 79.0 | 1193 | 1 | AX346077 | Trepnema |
| 250 | 15.8 | 79.0 | 1568 | 6 | AX577680 | Sequence |
| 251 | 15.8 | 79.0 | 1581 | 9 | AK025867 | Homo sapi |
| 252 | 15.8 | 79.0 | 1602 | 6 | AX577684 | Sequence |
| 253 | 15.8 | 79.0 | 1638 | 10 | BC028637 | Mus muscu |
| 254 | 15.8 | 79.0 | 1655 | 3 | CGY201070 | Coelogyne |
| 255 | 15.8 | 79.0 | 1692 | 9 | BC028014 | Homo sapi |
| 256 | 15.8 | 79.0 | 1703 | 3 | VS0270180 | Vannuccia |
| 257 | 15.8 | 79.0 | 1704 | 3 | CS0270168 | Calviria |
| 258 | 15.8 | 79.0 | 1749 | 9 | BC034321 | Homo sapi |
| 259 | 15.8 | 79.0 | 1877 | 9 | BC004526 | Homo sapi |
| 260 | 15.8 | 79.0 | 2064 | 6 | AX834571 | Sequence |
| 261 | 15.8 | 79.0 | 2064 | 9 | AK097109 | Homo sapi |
| 262 | 15.8 | 79.0 | 2431 | 1 | MSGBCG | |
| 263 | 15.8 | 79.0 | 2531 | 9 | HSNDSADH | |
| 264 | 15.8 | 79.0 | 2597 | 9 | HSAL7866 | H.sapiens m |
| 265 | 15.8 | 79.0 | 2689 | 10 | AK122329 | Mus muscu |
| 266 | 15.8 | 79.0 | 2731 | 6 | AX746759 | Sequence |
| 267 | 15.8 | 79.0 | 2731 | 9 | AK091062 | Homo sapi |
| 268 | 15.8 | 79.0 | 2747 | 6 | AK877398 | Sequence |
| 269 | 15.8 | 79.0 | 2747 | 6 | BD156637 | Primer fo |
| 270 | 15.8 | 79.0 | 2747 | 6 | AK001729 | Homo sapi |
| 271 | 15.8 | 79.0 | 2874 | 6 | AX880200 | Sequence |
| 272 | 15.8 | 79.0 | 2874 | 6 | BD158251 | Primer fo |
| 273 | 15.8 | 79.0 | 2874 | 9 | AK027636 | Homo sapi |
| 274 | 15.8 | 79.0 | 3396 | 8 | AK109980 | Oryza sat |
| 275 | 15.8 | 79.0 | 3926 | 9 | HSN801327 | Homo sapi |
| 276 | 15.8 | 79.0 | 4079 | 9 | AF037438 | Homo sapi |
| 277 | 15.8 | 79.0 | 4380 | 1 | MSGTCWPA | Mt.tuberculo |
| 278 | 15.8 | 79.0 | 4380 | 6 | I08847 | Sequence 5 |
| 279 | 15.8 | 79.0 | 4380 | 6 | I08856 | Sequence 10 |
| 280 | 15.8 | 79.0 | 4380 | 6 | AR262811 | Sequence |
| 281 | 15.8 | 79.0 | 4976 | 9 | AB009577 | Homo sapi |
| 282 | 15.8 | 79.0 | 4997 | 9 | HSN805783 | Homo sapi |
| 283 | 15.8 | 79.0 | 5054 | 9 | AB046853 | Homo sapi |
| 284 | 15.8 | 79.0 | 5103 | 9 | HSAA427355 | Homo sapi |
| 285 | 15.8 | 79.0 | 5339 | 10 | AK129480 | Mus muscu |
| 286 | 15.8 | 79.0 | 5499 | 9 | AB067511 | Homo sapi |
| 287 | 15.8 | 79.0 | 5914 | 9 | HSN805723 | Homo sapi |
| 288 | 15.8 | 79.0 | 5921 | 1 | AF222792 | Streptomy |
| 289 | 15.8 | 79.0 | 6003 | 9 | HSN805844 | Homo sapi |
| 290 | 15.8 | 79.0 | 6192 | 9 | AF448860 | Homo sapi |
| 291 | 15.8 | 79.0 | 6223 | 9 | AK122913 | Homo sapi |
| 292 | 15.8 | 79.0 | 6246 | 9 | AL833481 | Homo sapi |
| 293 | 15.8 | 79.0 | 7401 | 3 | DME309005 | Drosophil |
| 294 | 15.8 | 79.0 | 7614 | 3 | AY052072 | Drosophil |
| 295 | 15.8 | 79.0 | 10363 | 1 | AE005811 | Caulobact |
| 296 | 15.8 | 79.0 | 11588 | 1 | AE013213 | Thermoana |
| 297 | 15.8 | 79.0 | 12357 | 1 | AE002006 | Deinococc |
| 298 | 15.8 | 79.0 | 13045 | 1 | AE012171 | Xanthomon |
| 299 | 15.8 | 79.0 | 13268 | 6 | AX587577 | Sequence |
| 300 | 15.8 | 79.0 | 13348 | 3 | DME251802 | Drosophil |
| 301 | 15.8 | 79.0 | 13348 | 3 | DME251802 | Drosophil |

RESULT 1

AE007080

LOCUS

DEFINITION

complete genome.

ACCESSION

AE007080

VERSION

AE007080.1

KEYWORDS

GI:13882094

SOURCE

ORGANISM

Mycobacterium tuberculosis CDC1551

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex

1 (bases 1 to 17131)

REFERENCE

AUTHORS

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,

Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,

Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,

Gill,J., Mikula,A. and Bishai,W.

Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains

Unpublished

2 (bases 1 to 17131)

REFERENCE

AUTHORS

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,

Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,

Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,

Gill,J., Mikula,A. and Bishai,W.

Direct Submission

Submitted (25-APR-2001) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source

1. 17131

/organism="Mycobacterium tuberculosis CDC1551"

/mol_type="genomic DNA"

/strain="CDC1551"

/db_xref="taxon:83331"

/note="clinical strain"

104. 1357

/gene="Mt2390"

104. 1357

/gene="Mt2390"

/note="similar to SP:P71894.PID:1524272 GB:ALI123456;

identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="PE family protein"

/protein_id="AK4682.1"

/db_xref="GI:13882095"

/translation="MTANANLARKORHFTTINVRSSLDNNRSESRGTSMQELSVIPE

QVESAAQDLAGIRASIAAAAGTTAVSAADEVSTAIASIFGAYGRQCQLISA

QASAFHDEFVNLTKGTATYRNTFANAQSNVLNAVNPASRLIGHPSAAESVQNSAP

ALIGNMENTS

17131 bp DNA linear BCT 27-APR-2001

Mycobacterium tuberculosis CDC1551, section 166 of 280 of the

complete genome.

AE007080 AE000516

AE007080.1 GI:13882094

Mycobacterium tuberculosis CDC1551

Mycobacterium tuberculosis CDC1551

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex

1 (bases 1 to 17131)

REFERENCE

AUTHORS

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,

Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,

Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,

Gill,J., Mikula,A. and Bishai,W.

Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains

Unpublished

2 (bases 1 to 17131)

REFERENCE

AUTHORS

Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,

Kolony,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,

Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,

Gill,J., Mikula,A. and Bishai,W.

Direct Submission

Submitted (25-APR-2001) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source

1. 17131

/organism="Mycobacterium tuberculosis CDC1551"

/mol_type="genomic DNA"

/strain="CDC1551"

/db_xref="taxon:83331"

/note="clinical strain"

104. 1357

/gene="Mt2390"

104. 1357

/gene="Mt2390"

/note="similar to SP:P71894.PID:1524272 GB:ALI123456;

identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="PE family protein"

/protein_id="AK4682.1"

/db_xref="GI:13882095"

/translation="MTANANLARKORHFTTINVRSSLDNNRSESRGTSMQELSVIPE

QVESAAQDLAGIRASIAAAAGTTAVSAADEVSTAIASIFGAYGRQCQLISA

QASAFHDEFVNLTKGTATYRNTFANAQSNVLNAVNPASRLIGHPSAAESVQNSAP

TLGGHSTVTTAGLAQAQRAVATVEQQAATAVAPLPSAGAGLAQVTVNGVTTAGOGSAA
KLATALQSAWPLAKSGGEFIVAGOSALTGVALLOPAVVGVVQAGGTFLTAGTSAATG
LGLTLGAGVFSQGVNGLASGTAATGGLLGSAGVQLFSAFLLAVPTLGGVSL
AIATVQLVQGVHLSLVVNVVAVVAGIQAQTAGAQPAQGVNHTMLAAQILGAPGIAVLQT
AGGHPAQGGIHLTTAGNAATVVLIS"
complement (1392..2939)
/gene="MT2339"
complement (1392..2939)
/note="similar to PID:1813569; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="nitrite extrusion protein"
/protein_id="AAK46683.1"
/db_xref="GI:13882096"
/translation="MEQHTLLQREBSPSPAPSRRLRGGRHITHDPEDLGAWEAG
NKGIAARNLNSVTVTHLGSYVMTLPVLELLMQDVYGFSTSDKFLGTLTATLFGAF
LMPVALASAI FGGRNWATFSAILLIPALITGVLLTHPGLPLWFLYCAALTGLGG
NPASSMSNANAFPHRLKGSALGAGVNLGPVAILQVLALIAVTVGERKPYLVCL
YVVLVAIVGVSLFMNVQHRVQVNRRLPIVSALVSTRTDMLLSLLYLGFSGFIG
FHFVQGVQLTFLACGQSPARATLHVELAFVGPLLAARVAILYGRGLADNGSRLT
LIVFVMTLAAGLISASLEGRHVQHRGATWGVFVCFVALGAVRIYGLSLGNGSVYKMI
PTIFEASRLDLSAERDWSRIISGVVIGFAAFGALGGVGNNALESYLSFGSG
TDAFWIMCYAAAALVTWKVYDRRTVTDGMGLQALVQRPASTPAELIGRTOSDRF
SGCSISA"
complement (3173..3700)
/gene="MT2339"
complement (3173..3700)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46684.1"
/db_xref="GI:13882097"
/translation="MRRQSAVPIALALLALLALLIVLGLSGCAWKPTRSPPN
TKCSDGPTADVRAIAAVPIVPGSKWEITRHTNCRHLHWQIILPTIASQSTPQ
QLIFFDRNIPLAGSPTRNRPRTVPLPAGDDTVTVQYQMGQSDQECCTGIGTVRPHI
GSDGKLEALGSIPIHQ"
3574..4161
/gene="MT2339"
3574..4161
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46685.1"
/db_xref="GI:13882098"
/translation="WTVVLSAASRRNRPRPQSPVPTVPTVPESGRRYADARH
EISENWEFLFINAGVDPKARVPVPLQGRITPDVAGAIADDPMAARWI
GSRPCSLIGQNTMGDEMGLGFLAGQRCVDRLVNGASRSTCSRLPVIASVDERLSV
LKEVPRLHSISIFKGRPGEVLTVTGYNFRGP"
4586..6232
/gene="MT2339"
4586..6232
/note="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="malate oxidoreductase"
/protein_id="AAK46686.1"
/db_xref="GI:13882099"
/translation="MSDARVPRIPAAALSAPSNRNGVGFTHAQRRLGLTGRPLSAVLT
LQOARVWHLQSLATLGRNLLQLHYRHEVLYFKVLADHLPMLPVPVPTVTVGE
AIQRFDEVRQGLFLSIDEDIEEAFNTLGLGPEVDVLDIVCTDAEALIGDWGV
GGIQAIVGKALYTAGGVDPDRCLAVSDVGTNDNEQLLADPFYLGNRHARRGREYD
EFVSRYLEAQTALFFRAILHFFDFGPNARKILDTYDTCVFNDDMOGTQAVLAAV
YSLKVTGTPIRQTVPGACAGNADQIRDAWADGATLEQAVSQIWPIDRGL
LFDMDMLDRDQFQPYAKNRHQLGVAGVDRVGLSDAIKIASPTILIGCSTVYGAFTKEV
VEANTASCKHPMIFPLSNPTSRMEALPADVLAWNSGRALLATGSPVAVPEFDETTIYVI

GOANNVLAPFGIGLIVAGARLITRMLHAAAKAIAHQANPTNPGDSILLPDVQNLRA
ISTTVAAVYRAVQGVASRTHDDVYRQAIVDITWMLPAYD"
complement (6186..7799)
/gene="MT2395"
complement (6186..7799)
/note="similar to GP:5881853; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="drug transporter"
/protein_id="AAK46687.1"
/db_xref="GI:13882100"
/translation="NNRTQLTLTIATGLFMIFLDALINVVALPDQRSFAVGEDGL
QWVVASVLSQMAVFIMSAATLADLYGRRRWYLGVSFLTGLSGIACGLAPSIAVLITAR
GAOGLGAAAVSVTSLALVSAAPPEAKERAKAIGITWAIASIGTTTPTLGGLLVDMQW
WRSIFYVNLPMGALVLFLLCYVEESCNERRARFPLSGOLLFTVAVGALVAVIEGPQ
IGWTSQTVLMTAAVGCALFVWLERRSNPMMLTLRDTSYALAIATICTVFEAV
YGMLLTTOFLQNVRGYTFSTVGLMILPFSAAVIVSPVHLNGRIGRARGVPILAGLC
MLMLGLMLTIFSEHRSALVGLGCGSVALCLPTITVAVTAPAEARASGIM
SAQRAICSTIGFAVLGSLAAMLSATLEPHLEAVDPVQHVLABIIIDSANPRAHV
GGIVPRRHIEHDPVAIAEDFTGIRVALLVATATLAVVFLAGMFWFRDVTHTAGSD
LSRLPTAMTVECAVSHMPGATWCRWPA"
7826..7996
/gene="MT2396"
7826..7996
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46688.1"
/db_xref="GI:13882101"
/translation="MAICDQTLGCSAGSVSRFGRFEPFNGRRLNRPNQWGRKFTAGVS
QLGEPITLAHGL"
8274..9206
/gene="MT2397"
8274..9206
/note="similar to PID:2342631; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cysteine synthase"
/protein_id="AAK46689.1"
/db_xref="GI:13882102"
/translation="MSIAEDITOLIGRTPLVRLRVRTDGAVIDIVAKLEFFNPANSVK
DRIGVAMLOAQEQGLIKPDTIILEPTSGNTGIALVCAARGYRCLVTFPMTSLER
RMLIRAYGAELITPGADGMSGAIAKAEELAKTDQRYFVFPQFPANPAIHRVTAE
EVRDRTDKVDIVVAGVGTGTTVAQVTKERKSFARFVAVPEASPVLSGQKGFH
PIQIGAGFVPVLDODLDEIITVGNEDALNVARRLAREBGLLVIGISSAATAVAALQ
VARRPENAGLIVVLPDFGERVLSPLFADVAD"
9210..9899
/gene="MT2398"
9210..9899
/note="similar to GP:2342632; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine acetyltransferase CysB, putative"
/protein_id="AAK46690.1"
/db_xref="GI:13882103"
/translation="MLTAMRGDIRAAREDPAAPTALEVIFCYGVHVGHRHLAHL
WQGARLARAARAEFTRLTGVDIHPGAVIGARVFIHATGVVIGTAEVGDVITYH
GVTLGSGVQKGRHPTVGDVRIIGAGAKVIGPIKIGEDSRICANAVVVKVPPSAV
VGVPQVIGVQSQSPGPPDWRLPDLVGASLDSLLTRVARELDALGCGPQAAGVIRPPE
AGIWHGEDFSI"
10315..11283
/gene="MT2399"
10315..11283
/note="identified by Glimmer2; putative"


```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AK46691.1"

Query Match          95.0%; Score 19; DB 1; Length 17131;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2  GAGGGCCCGCCAGCCACAA 20
      |||||
DB      3615 GAGGGCCCGCCAGCCACAA 3633

RESULT 2
BX248342
LOCUS   Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
DEFINITION
ACCESSION BX248342 BX248333
VERSION   BX248342.1 GI:31619031
KEYWORDS  complete genome.
SOURCE    Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM  Mycobacterium bovis subsp. bovis AF2122/97
          Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
          Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
          tuberculosis complex.
REFERENCE
AUTHORS  Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
          Pryor,M., Duthey,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
          Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
          Whesler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
          Hewinson,G.
          The complete genome sequence of Mycobacterium bovis
          Online Publication
          PNAS 10.1073/pnas.1130426100 ( Microbiology )
          2 (bases 1 to 306550)
          Garnier,T.
          Direct Submission
          Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
          Bacterienne Institut Pasteur,28,rue du Dr Roux 75724 PARIS cedex
          15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
          Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
          Laboratories Agency Weybridge, Woodham Lane, New Haw Addlestone,
          Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
          Hinxton, Cambridge CB10 1SA, UK. PH4 Annotation, Gencopie, Institut
          Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
          Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
          rue du Docteur Roux, 75724 Paris Cedex 15, France
          Location/Qualifiers
          1. 306550
             /organism="Mycobacterium bovis subsp. bovis AF2122/97"
             /mol_type="genomic DNA"
             /strain="AF2122/97"
             /db_xref="taxon:233413"
             201..1286
             /gene="adhE2"
             /locus_tag="Mb2283"
             201..1286
             /gene="adhE2"
             /locus_tag="Mb2283"
             /EC_number="1.2.1.-"
             /note="Mb2283, adhE2, len: 361 aa. Equivalent to Rv2259,
             len: 361 aa, from Mycobacterium tuberculosis strain H37Rv,
             (100.0% identity in 361 aa overlap). Probable adhE2,
             zinc-containing alcohol dehydrogenase, similar to several,
             especially mycothiol-dependent formaldehyde dehydrogenase
             from Amycolatopsis methanolica P80094 (360 aa). Contains
             P500059 Zinc-containing alcohol dehydrogenases signature.
             FASTA scores: >sp|P80094|FADH AMYME
             NAD/MYCOTHOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE
             (MD-FALDH) Length = 360, Expect = e-156, Identities =
             268/358 (74%). Also similar to Rv0162c, (MTC128.02c, 35.0%
```

```
identity in 371 aa overlap)."
/codon_start=1
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase
adhE2"
/protein_id="CAD97144.1"
/db_xref="GI:31619032"
/transl_table="MSQTVRGVIAQKGFVELVNIWVPDPGGEAVVDVTACGVRG
DLVREGGINDEYFPLIGHEAGII EAVGVPTAVPGDFVILNWRVACQCRACKRG
RPRVPTFNAEOKMTLTDGTELTALGAFADKTLVHSGQCTKYDPAADPAVAGLL
GCGMAGLGNAINTGVTTRDDTVAVIGCGVGDAIAGALVGAKRIIAVDTDDTKLD
WARTFGATHIVNAREVDVVOAIGLIDGFGADVVIDAVGRPEFYQQAFTAYRDLAGIVV
LVGVPDMDRLMPLVDFSHGGLGSGWYDCLPESDFFTLIDLVLQGLRPLQRFVVS
ERIGLEDVEEAFPHKMGKVLRSVVM"
1286..1921
/locus_tag="Mb2284"
1286..1921
/notes="Mb2284, -, len: 211 aa. Equivalent to Rv2260, len:
211 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 211 aa overlap). Conserved
hypothetical protein, similar to hypothetical proteins
Rv0634c, Rv1637c, Rv3677c, Rv2581c from Mycobacterium
tuberculosis and to various hydrolases. FASTA scores:
spr|006154|006154 HYPOTHETICAL 21.3 KD PROTEIN (200 aa)
opt: 355, E(): 4e-15; (37.4% identity in 198 aa
overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97145.1"
/db_xref="GI:31619033"
/transl_table="MAAERVITHGTFELDGGSWEDNNIWLVDSEVVVFPAAHHA
APIIDAVGGRKVVAVICTGHNDHVTVAPELGTALDAPVLMHFGDVALMRTHPDKSF
RVIDGDGAVRGVTELRALHTPGHSPGVCWYAPELGPGTGTGVTGDTLFAAGFGATG
RSYSDPFTILRSISGRIGALPGDVTVHTGHDSTTICDEIVHYEEVWARGH"
complement(1998..3506)
/locus_tag="Mb2285c"
complement(1998..3506)
/notes="Mb2285c, -, len: 502 aa. Equivalent to Rv2262c and
Rv2261c, len: 360 aa and 140 aa, from Mycobacterium
tuberculosis strain H37Rv, (94.7% identity in 357 aa
overlap and 100.0% identity in 140 aa overlap). Conserved
hypothetical protein, with function unknown but some
similarity to N-terminal 70% of
P23930|P77703|LMT_ECOLI|CUTE|E0657 APOLIPOPROTEIN
N-ACYLTRANSFERASE [EC 2.3.1.-] from Escherichia coli
strain K12 (512 aa), FASTA scores: opt: 239, E(): 1.6e-07,
(30.4% identity in 359 aa overlap). Note that neighboring
ORF shows similarity to N-terminal part of PCC6803
apolipoprotein N-acyltransferase from Synechocystis sp.,
suggesting possibility of frameshift. Sequence of clones
from two sources has been checked but no error found.
Appear to be two extra bases at position 1876970 compared
to CDC1551 strain. Conserved hypothetical protein, with
function unknown but some similarity to C-terminal end of
PCC6803 apolipoprotein N-acyltransferase from
Synechocystis sp. Note that next ORF shows similarity to
N-terminal part of P74055 APOLIPOPROTEIN N-ACYLTRANSFERASE
from Escherichia coli (519 aa), FASTA scores: opt: 142,
E(): 0.007, (29.9% identity in 117 aa overlap), suggesting
possible frameshift. Sequence of clones from two sources
has been checked but no error found.
REMARK-M.bovis-M.tuberculosis: In Mycobacterium
tuberculosis strain H37Rv, Rv2262c and Rv2261c exist as 2
genes. In Mycobacterium bovis, a 2 bp deletion (ct-*)
results in a single product which is more similar to
Rv2262c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97146.1"
```



```

RESULT 4
RMU22370/c      1073 bp  DNA      linear      BCT 24-NOV-1995
LOCUS           Rhizobium meliloti insertion sequence ISRM2011-2 putative
DEFINITION      transposase and ORF genes, complete cds.
ACCESSION       U22370
VERSION         U22370.1  GI:1072026
KEYWORDS
SOURCE          Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM        Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE       1 (bases 1 to 1073)
AUTHORS         Selbitschka,W., Arnold,W., Jording,D., Kosier,B., Toro,N. and
                Puhler,A.
TITLE           The insertion sequence element ISRM2011-2 belongs to the IS630-Tc1
                family of transposable elements and is abundant in Rhizobium
                meliloti
JOURNAL         Gene 163 (1), 59-64 (1995)
MEDLINE         96001244
PUBMED          7557479
REFERENCE       2 (bases 1 to 1073)
AUTHORS         Selbitschka,W.
DIRECT SUBMISSION
TITLE           Submitted (08-MAR-1995) Werner Selbitschka, Dept. of Genetics,
                University of Bielefeld, Universitaetsstr. 25, Bielefeld, NRW 33501,
                Germany
FEATURES         Location/Qualifiers
source           1..1073
                /organism="Sinorhizobium meliloti"
                /mol_type="genomic DNA"
                /strain="2011"
                /db_xref="taxon:382"
source           11..1063
                /organism="Sinorhizobium meliloti"
                /mol_type="genomic DNA"
                /strain="2011"
                /db_xref="taxon:382"
                /insertion_seq="ISRM2011-2"
misc_feature     1..10
                /notes="part of E. coli rpsL gene"
misc_feature     9..10
                /notes="target duplication"
repeat_unit     11..29
                /rpt_type="inverted"
RBS             89..94
CDS             join(104..451,451..1050)
                /function="mediates transposition"
                /notes="putative transposase"
                /codon_start=1
                /transl_table=11
                /protein_id="AAC43490.1"
                /db_xref="GI:1072027"
                /translations="MARFSDLRERVDAVTGEGLSCEAAAKRRGIGISTADWVR
                FRGTGAAPQMGHPRKLSGHRALWLCRCRERDFTLHGLVASERSGKLKVDYRAV
                WTVFHEGLSYKRTIVASEREPDVARHARWLKPCPIDFARLVFIDETWTKNNA
                PLRWGAPRGRLVGYAPFGHWTMTFVAALRADRVSAFFILDPGPIRGFRFRIYVOQVL
                VPELKAGDIVLNLGSHKGQETIRAAIRKAGARLFLPKYSPDLNPILKFAKIKHL
                REAQRASDAIHDLRHILQAVTPQCEAAVFEAGYERA"
                complement(439..810)
                /functions="unknown"
                /notes="ORF"
                /codon_start=1
                /transl_table=11
                /protein_id="AAC43491.1"
                /db_xref="GI:1072028"
                /translations="MGARIVQDDVAGFEFRHQNLLIDAAFAIDRAIEDKGSADAV
                GPFCGKGHGVPNAEGVAHQSFARRPAPQRHVRPLRLGLIDENAGGINSGLVLP
                SCPVAGDVGLPLFAGDQRULF"
misc_feature     443..450
                /notes="putative ribosomal slippage site"
stem_loop       452..466
                /notes="inverted repeat"

```

```

/function="putative involvement in ribosomal
frameshifting"
repeat_unit     1045..1063
                /rpt_type="inverted"
misc_feature     1064..1073
                /note="part of E. coli rpsL gene"
misc_feature     1084..1085
                /note="target duplication"

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 1073;
Best Local Similarity 94.7%; Pred. No. 9.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CGAGGCGCCCGAGCCGCCACA 19
        |||||
Db      605 CGAGGCGCCCGAGCCGCCACA 587

RESULT 5
AK098077/c
LOCUS           AK098077      2253 bp      mRNA      linear      PRI 15-JUL-2002
DEFINITION      Homo sapiens cDNA FLJ40758 fis, clone TRACH2002333, highly similar
                to Human phosphotyrosine independent ligand p62 for the Lck SH2
                domain mRNA.
ACCESSION       AK098077
VERSION         AK098077.1  GI:21758012
KEYWORDS        oligo capping; fis (full insert sequence).
SOURCE          Homo sapiens
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1
AUTHORS         Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
                Arita,M., Musashino,K., Iuuki,H., Hara,H., Sugiyama,T., Irie,R.,
                Osuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
                Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
                Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
                Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
                Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K.
                and Isogai,T.
                NEDO human cDNA sequencing project
                Unpublished
                2 (bases 1 to 2253)
                Isogai,T. and Yamamoto,J.
                Direct Submission
                Submitted (04-JUL-2002) Takao Isogai, FLJ project (HRI Team); 2-6-7
                Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
                (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
                NEDO human cDNA sequencing project supported by Ministry of
                Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                Research Association for Biotechnology (RAB); cDNA library
                construction: Helix Research Institute (HRI) (supported by Japan
                Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
                HRI, and Biotechnology Center, National Institute of Technology and
                Evaluation; clone selection for full insert sequencing: HRI and
                RAB; annotation: HRI and RAB.
                Location/Qualifiers
FEATURES         source
                1..2253
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="TRACH2002333"
                /tissue_type="trachea"
                /clone_lib="TRACH2"
                /note="cloning vector: pME18SFL3"

ORIGIN
Query Match      87.0%; Score 17.4; DB 9; Length 2253;
Best Local Similarity 94.7%; Pred. No. 9.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GAGGCGCCCGAGCCGCCACA 20

```

```

Db          89 GAGGGGCCAGCCCCAACAA 71
||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
AE007196/c
LOCUS      Sinorhizobium meliloti plasmid pSymA section 2 of 121 of the linear BCT 15-AUG-2001
DEFINITION complete plasmid sequence.
ACCESSION AE007196 AB006469
VERSION    AE007196.1 GI:14523063
KEYWORDS
SOURCE     Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM   Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE 1 (bases 1 to 10029)
AUTHORS    Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubier,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kallman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSymA megaplasmid
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
JOURNAL    21396509
MEDLINE    11481432
PUBMED
TITLE      2 (bases 1 to 10029)
AUTHORS     Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubier,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kallman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES   Location/Qualifiers
            source           1..10029
                        /organism="Sinorhizobium meliloti"
                        /mol_type="genomic DNA"
                        /strain="1021"
                        /db_xref="taxon:382"
gene        158..715
                /plasmid="pSymA"
                /gene="SMA0017"
                /gene="SMA0017"
                /function="Miscellaneous; Unknown"
                /notes="glimmer prediction"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK64666.1"
                /db_xref="GI:14523064"
                /translation="MAADDDHGCGGLRQKARGLPOARNHLSGTRICKRQAQIIL
                GAGGPGNEPGRCCMPGREIPEDGGENSRKRSAWGRRQXTSLTAFIAELIR
                ANVEERLTPYISRLSDVDTIEDMRQTGAASHARDVDWLQLASARADLSA
                ASTRDVLDAADIIRTKIVLDGKE"
repeat_region 1310..2368
                /standard_name="ISRm2011-2/ISRm11"
                /note="SM3000; predicted by homology"
                /rpt_family="ISRm2011-2/ISRm11"
                /length=1067
                /gb_xref="GI:14523064"
                /gene="SMA0018"
                /gene="SMA0018"
                /function="Elements of external origin; Transposon-related
                functions"
                /notes="glimmer prediction"
                /codon_start=1
                /transl_table=11
                /product="TRM2011-2a transposase"
                /protein_id="AAK64667.1"

gene        158..715
                /plasmid="pSymA"
                /gene="SMA0017"
                /gene="SMA0017"
                /function="Miscellaneous; Unknown"
                /notes="glimmer prediction"
                /codon_start=1
                /transl_table=11
                /product="TRM2011-2a transposase"
                /protein_id="AAK64667.1"

/db_xref="GI:14523065"
/translation="MARFNSDLRVRVAVTGEGSCRAAKRFIGIGISTADTWDRVA
FRETSAAPQGGMGHKPKLGSPHRAWLLCRCDERDFTLHLVAELSRLGVKYDYRA
WTVEEGLSYKKRWSPANGSGPTSPATCHDG"
gene        1879..2352
                /gene="SMA0020"
                /gene="SMA0020"
                /functions="Elements of external origin; Transposon-related
                functions"
                /note="glimmer prediction"
                /codon_start=1
                /transl_table=11
                /product="TRM2011-2b transposase"
                /protein_id="AAK64668.1"
                /db_xref="GI:14523066"
                /translation="MAPLRGWAPGERLGVYAPEFGHWNTMTFVAALRADRVSAFFILD
                GPINGERFRIYVOOVLPELVKAGDIVILNLGSHKGGEIRAAIKKAGARLFPLPKYPX
                DLNPTEKFLFAKIHKWLREAQRSRDAlHDELRLILOAVTQECAAAYKEAGYERA"
gene        3102..3746
                /gene="SMA0021"
                /gene="SMA0021"
                /functions="Elements of external origin; Transposon-related
                functions"
                /note="glimmer prediction, with response regulator
                receiver domain"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK64669.1"
                /db_xref="GI:14523067"
                /translation="MYRLYNHAIRFSLDCPEGLTGDKTRLKAQAMLIAELHOLGYR
                DMPQRCAYRGLESYEKIAKPRSTSCSEADMKPATILLAREALLLDLYEA
                LDAGFEVVAVARGKALEWRSDSVAGVATDIRVELPNHGHSVARARETYPGIF
                IVYGGHGLEAWRSRGVANSILLEKPFALAQLVTAVSELLNEPVLLSVADPDNP"
gene        4368..4459
                /gene="selC"
                /note="synonym: SMA0022; selenocysteine tRNA in FdH
                cluster"
                /protein_id="AAK64670.1"
                /db_xref="GI:14523068"
                /translation="MVVILSERAFSRSDQTGAEVMLNRFRFLTNAATAAATLVWSGS
                GMQARAEEAGNGSPWRFEITTVKDLPAGEPAQLWLPVVSAGDYQKADAPLVWSNS
                TDIREKDPSSGADILRMWEDEPRRVIEVTHVATDRRTVKSIATAEALQHELGR
                TSPMPVDGIVKTAMRIVEGDRAEDRARAIYDWINDTFRFDANIICGGVGNARDMLE
                TCYFGKCADITSSLFVSLARAAGLPARDVFGRVADSADFKSLGRSGDIKAQHCRAE
                VYLDAHGWVPDVPADVRRKVLEENLPLDNPAVRAPREKAGNVWNWVGINTARDLV
                PGGEHQGLFMYPAAVTSRGCCLNPPQTAYSITSREITA"
                /complement(6076..6504)
                /gene="SMA0026"
                /gene="SMA0026"
                /complement(6076..6504)
                /function="Miscellaneous; Unknown"
                /note="glimmer prediction"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"

```

[illegible]

```
/protein_id="AAK64671.1"
/db_xref="GI:14523069"
/translation="MEQAPFNIVGGDEKSIAMNLAFTPSACVLAARFLIALANP
ASAQNTNEDCAAGAPATPPDRASADGTAPCNSGSGTGTGCGAHGNTNPGQ
ATAGSTTWPPTTARGIDLATAPETAPAPETAPAGTQTPTDC"
gene
complement (6517..7566)
/gene="seld"
/gene="seld"
/notes="synonym: Sma0028"
complement (6517..7566)
/gene="seld"
/EC_number="2.7.9.3"
/function="Small molecule metabolism; Global functions"
/notes="glimmer prediction"
/codon_start=1
/transl_table=11
/product="selenide, water dikinase (selenophosphate
synthetase), Seld"
/protein_id="AAK64672.1"
/db_xref="GI:14523070"
/translation="WNTLAPRLTDLAHGGCGCKLAPSLQOLLANPAARFPAQLLV
GNDMADDAWQVDDNTCVIATDTPMPVDDPRDRIASNAISDVYANGKPLA
LAIVGPNKJLSTIAKLEGASICAERAGIPVAGCHSIDSVETIYGLAIVGICHP
EVRNGGKAGDALILTGGVIGYSAIKKALPPGCVCEMIGSTILLNIGADLAA
DPDVALTDTGVLGHLEMARASQLTLRLSAIPFLSQAYMLAQFGITGASGR
NWASYGADVLPDLPDMQRLLLADPQTSGLLVACAPKAGALVEKVLGAYPLAGI
VGTAAAGAAQIKVIS"
gene
complement (7861..8088)
/gene="Sma0031"
complement (7861..8088)
/gene="Sma0031"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK64673.1"
/db_xref="GI:14523071"
/translation="MSEWIDFRWDPDKMERPGIVFEVTNGDQTLTGCVVPLPLPS
DWVHPLRPRAVQPRRHSSPLPKPAGPQQ"
gene
complement (8105..9760)
/gene="Sma0034"
complement (8105..9760)
/gene="Sma0034"
/EC_number="3.4.24.40"
/function="Macromolecule metabolism; Macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="glimmer prediction; partially similar to serralyisin
(serratia protease) precursor"
/codon_start=1
/transl_table=11
/product="possible protease"
/protein_id="AAK64674.1"
/db_xref="GI:14523072"
/translation="WPAATYATATGNAYIDGLGDKWGIKIDTFSPFPTSASFSAGY
GNGEPLKGPAAALSGQAATRALQDFSSVANVTETESATKHADLRASSDAPST
AWAYFSTAEAGDAWFKSSGHYSRPVKNYAYVFLHETGHALGLHEAGNMPV
NRDSMEYTVMSYRSYVGAATTTGYNETWGYAQLMMYDIALQHYMGADFTTSENT
TYRSPSTGEMFVNGVGGAGPGCKILLTVWDGGDTDYFSNYTTALKVLDLRPEWT
TTSAQLAKLHYDGSKVAIGNALQYQGDTRSLIENAKGAGNDAITGNAANALW
```

Query Match 87.0%; Score 17.4; DB 1; Length 10029;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCGCCA 19
|||||
Db 1907 CGAGGCGCCAGCCGCCA 1889
|||||

RESULT 7
AE007297/c
LOCUS

```
DEFINITION Sinorhizobium meliloti plasmid pSma section 103 of 121 of the
complete plasmid sequence.
ACCESSION AE007297 AE006469
VERSION AE007297.1 GI:14524266
KEYWORDS
SOURCE
ORGANISM Sinorhizobium meliloti (Rhizobium meliloti)
REFERENCE
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hymen,R.W., Kahn,D., Kahn,M.L.,
Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSma megaplasmid
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
21396509
11481432
2 (bases 1 to 10296)
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hymen,R.W., Kahn,D., Kahn,M.L.,
Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
Location/Qualifiers
1..10296
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSma"
164..703
/gene="Sma2029"
164..703
/gene="Sma2029"
/function="Small molecule metabolism"
/notes="glimmer prediction. Contains similarity to
Isochorismatase family. Similar to LOCUS 4558196 208 aa
Chain A, The 1.8 Angstrom Crystal Structure Of The Yeac
Gene Product From Escherichia Coli Reveals An Octameric
Hydrolase Of Unknown Specificity. ACCESSION 4558196"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65768.1"
/db_xref="GI:14524267"
/translation="MVNHTALAKLAPAFNVPTILTSVIAARGLLFKQITDVPDQ
ASGCIKSGEHAIVQIRMAAGANVTVMALAGEWORDNARTEHVELTILIQHFGS
GIAYLEOOLLNTPVPEEG"
714..1085
/gene="Sma2031"
714..1085
/gene="Sma2031"
/function="Cell Processes; Protection responses;
detoxification"
/notes="glimmer prediction. Local similarity to 278 aa
NON-HEME CHLOROPEROXIDASE. Possible protein fragment."
/codon_start=1
/transl_table=11
/product="Putative non-heme chloroperoxidase"
/protein_id="AAK65769.1"
/db_xref="GI:14524268"
/translation="MAKAVLVAAIPPLMLKNDDDPGTPEVDFGFRFALAGNRAOFF
RDVPSGFGFSGFNREGAAVHEGVQLNWRQGMKRGALRWQGLLGNRPRTSRRS
PCQRRCCTVTRTSFRSPRL"
923..1222
/gene="Sma2032"
```

```

CDS
923..1222
/gene="Sma2032"
/function="Cell Processes; Protection responses; detoxification"
/note="glimmer prediction. Local similarity to 278 aa NON-HEME CHLOROPEROXIDASE. Possible protein fragment."
/codon_start=1
/transl_table=11
/product="Putative non-heme chloroperoxidase"
/protein_id="AAK65770.1"
/db_xref="GI:14524269"
/transl_table="MAAGHDGEAPRRITMASRPRKPTKTDLKAITVPTFVLHGEDD
OVPIAAGALKVLLPNSGLKTPFGSHGLMTVNADVLNADLLEFTRSDCGA"
1755..2552
/gene="Sma2033"
1755..2552
/gene="Sma2033"
/function="Miscellaneous; Unknown"
/note="glimmer prediction. C-terminus is similar to LOCUS BAA19632 376 aa lytic transglycosylase [Actinobacillus actinomycetemcomitans]. ACCESSION BAA19632"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65771.1"
/db_xref="GI:14524270"
/transl_table="MAFFAKQLIAAGSLAPCLITCAVEAASCGGAAGFOQWKVEF
AETAKAGVRGGLAALIGAXYVGTIRADRSVNAKAFSGYVDDFWRRGSAIISKGR
SLKTAALAKNIERKIDVPAGVLLAIGMEITGFGASMGNTVSAITVLYADCRPQ
FAPHAIALKLVDSGLSVARSVGAHGEIGOTGPLGNVLKIVGVSRNMDTSTALM
STANFLKAHGRAGGYQGNMGAAGNSASVYQKAIARIGEAIDGR"
complement(3483..3728)
/gene="Sma2034"
complement(3483..3728)
/gene="Sma2034"
/function="Miscellaneous; Unknown"
/note="glimmer prediction. Similar to C-terminus of LOCUS BAA19632 376 aa lytic transglycosylase [Actinobacillus actinomycetemcomitans]. ACCESSION BAA19632"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65772.1"
/db_xref="GI:14524271"
/transl_table="MVEVTLWGLAAAGNSKVEIAKNIRELFARLSERFPRLEPL
MARGTAVADGTIVRDTWSKELPTGAEIYLLPRLAGG"
complement(3738..6170)
/gene="Sma2037"
complement(3738..6170)
/gene="Sma2037"
/function="Small molecule metabolism"
/note="glimmer prediction. Similar to motifs found in aldehyde oxidase and xanthine dehydrogenase, C terminus. Similar to LOCUS CAA05038 769 aa 4-Hydroxybenzoyl-CoA reductase alpha-subunit [Thaurea aromatica]. ACCESSION CAA05038"
/codon_start=1
/transl_table=11
/product="Putative oxidoreductase"
/protein_id="AAK65773.1"
/db_xref="GI:14524272"
/transl_table="MWHLTRPCHRCKSVAEESGSDRRGNTLRCRKSLSIHLRQD
RARRGRSRHRSHTMNFDPHRSFSSVGTFRIPDPGVKVTGRARYGADFNMPG
QIVGRILRSFHAHAIIRKIDTLKAAQLPGVAVLPAADLPDLDGDAAMYDILNCMA
TKKALYDGHAVAAVNAARTAKALIEVEYELLPHVTDVQAFADPAIDNCTIF
TTGVDPKPPRNSMRSQFSGHDVDMARADVFVETFKTEQHOGYIPHACVAS
VSDGTADLVCTQGHFVYRCHCAOLLGMDASKLAVTSSEIGGFGGKTHVWAEVAL
ALSRAGRPKVLMTEDEVFRASGTSATSIDVKIGALKDGTIIAADATLAYSQGPYA
GAWEAVGMAFACVLENTVTVGEVLNREKTAAYRAPAPMAFAVEASIDELAX
KYGMNPFRIENAAQEGKASGYPIGIGITPLEAVKNHPHMKAPLGRNQGGMA
CGFWNFQGTCTDLNIGMDGTVSLAVGTVDVGGSRASLSLVAAEBELGIDYSQRTVI

```

ADTSLGYNDTDSRGTFSSSWATISAARNAIKILRRAQMWDPVEDVTWKGHA
 VAKGETHNLKLSLKEIAAASGTTGGPIAGHSEIVADGAGVSEFATHICDIEVDPEIG
 ATKLYRYTVQDAGKAVHETTYEGQYCGGAAGIGWALNEEYVYNGDGLQAGSLDY
 RIPCSDLPMDITQILEIFNPNHPYGVGVGETSIVPPLAAVANAVSNVAGVRLHIP
 MSPRILAALEA
 complement(6469..7329)
 /gene="Sma2041"
 complement(6469..7329)
 /gene="Sma2041"
 /function="Small molecule metabolism"
 /notes="glimmer prediction"
 /codon_start=1
 /transl_table=11
 /product="probable oxidoreductase"
 /protein_id="AAK65774.1"
 /db_xref="GI:14524273"
 /transl_table="MRYIRPSIEDAVGLLAASGKAVALAGSDLLVRMKGFIPE
 LIIDIKALDRHITESDEGFVIGAAVPCAVIGENALARRPVGVEGANLGSQKQYQ
 GRTIVNJCNASPAADSVFALLAGALIRPGAGSRITAVEMVPGVKISLAPGE
 IIEAILDKRQRPDAYQRTPTREMDIAVVAAYKLTLDQGVQVSARVALGAAP
 TVLLVEAAEILTGSVRDDKTLDRLLAAGSGACRPIDDKRGTVFERRKVAGVLAKRVA
 LAAYQRAGLE"
 7882..8941
 /standard_name="ISRM2011/ISRM11"
 /notes="Sma3016; predicted by homology"
 /evidence="not experimental"
 /rpt_family="ISRM2011-2/ ISRM11"
 7977..8387
 /gene="Sma2043"
 7977..8387
 /gene="Sma2043"
 /functions="Elements of external origin; Transposon-related functions"
 /codon_start=1
 /transl_table=11
 /product="TRM2011-2a transposase"
 /protein_id="AAK65775.1"
 /db_xref="GI:14524274"
 /transl_table="MARPSNDLRRVAVDVTGEGLSGSCRAAAKRFQHRHQRDLGAA
 VSRDQRRTPDQMAQAPQAFRAGLALPLPRALHAARTCRRVERARPEGGFIGA
 VMTFVHEEGLSYKKRRWSPANGSGTSPATHDG"
 8453..8926
 /gene="Sma2045"
 8453..8926
 /gene="Sma2045"
 /functions="Elements of external origin; Transposon-related functions"
 /codon_start=1
 /transl_table=11
 /product="TRM2011-2b transposase"
 /protein_id="AAK65776.1"
 /db_xref="GI:14524275"
 /transl_table="MAPLRGWAPRGLVGYAFPGHWNMTTFVAALRADRVSAFFILD
 Query Match 87.0%; Score 17.4; DB 1; Length 10296;
 Best Local Similarity 94.7%; Pred. No. 1e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 CGAGGCGCCAGCCCA 19
 |||||
 Db 8481 CGAGGCGCCAGCCCA 8463
 RESULT 8
 AE007314/c
 LOCUS AE007314 10578 bp DNA linear BCT 15-AUG-2001
 DEFINITION Sinorhizobium meliloti plasmid pSymA section 120 of 121 of the
 complete plasmid sequence.
 ACCESSION AE007314 AE006469
 VERSION AE007314.1 GI:14524456
 KEYWORDS Sinorhizobium meliloti (Rhizobium meliloti)
 SOURCE Sinorhizobium meliloti
 ORGANISM Sinorhizobium meliloti

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

REFERENCE 1 (bases 1 to 10578)
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.

TITLE Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)

MEDLINE 21396509

PUBMED 11481432

REFERENCE 2 (bases 1 to 10578)
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2001) Biological Sciences, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA

FEATURES
 source Location/Qualifiers
 1..10578 /organism="Sinorhizobium meliloti"
 /mol_type="genomic DNA"
 /strain="1021"
 /db_xref="taxon:382"
 /plasmid="pSymA"
 420..1478 /standard_name="ISRM2011-2/ISRM11"
 /notes="Sma3018; predicted by homology"
 /evidence="not experimental"
 /rpt_family="ISRM2011-2/ISRM11"
 516..923 /gene="Sma2373"
 516..923 /gene="Sma2373"
 /codon_start=1
 /transl_table=11
 /product="Tm2011-2a transposase"
 /protein_id="AAK65941.1"
 /db_xref="GI:14524457"
 /translation="MARFSDNRERVDVAVTGEGLSCRAAKRFGIGISTADWRR
 FRFTGAAPQMGKPKRLSGPHAWLRCRERDFTLHGLVAELSERGLKVDYRAV
 WTVFDEGSLYKRRWSPANGSGPTSPATGHDG"
 989..1462 /gene="Sma2375"
 989..1462 /gene="Sma2375"
 /functions="Elements of external origin; Transposon-related functions"
 /codon_start=1
 /transl_table=11
 /product="Tm2011-2b transposase"
 /protein_id="AAK65942.1"
 /db_xref="GI:14524458"
 /translation="WAPLKGWAPRGERLVGYPGCHWNTWTFVVALRADRVSAFILD
 GPINGFRFRIYQVQVLVPELKGADIVLDNLSHGKQEIARAIRKAGARLFLPKRSP
 DLPNFKLPAKTKHMLREQAQRSDAIHDELHILQAVTPQCAAYFKEAGYERA"
 1664..2833 /gene="Sma2377"
 1664..2833 /gene="Sma2377"
 /functions="Cell processes; Transport of small molecules"
 /notes="glimmer prediction; similar to putative chloramphenicol transporter in several organisms, including Salmonella typhimurium, AF118107, Klebsiella pneumoniae, AAF27726, and Enterobacter aerogenes, AAD22144"
 /codon_start=1
 /transl_table=11
 /product="putative transmembrane transport protein"

/protein_id="AAK65943.1"
 /db_xref="GI:14524459"
 /translation="MTVNVSPVTVMFLISFTVGLIAGTDLVLPALPAMPALTGGTAA
 LAQNVLAAYAGGLTVGLITFEGELGYSRRKLLVSLGFAVTSLSASVAPTLEWVVI
 LRPAQAFSGSPAVFAPFGTHGLYPGDKAPSPFMGRGCSIESLPALPAGATVMTVG
 GWTSFLMLAGLAILCAVGSWAYRQSLPDRLEALEVHQSVMSIIRNGDFLRHGLSOAL
 SLGTSILFVGPAPAVMTGALMTGTFILLOVFGIALFILASNASNALPARRFGIERMI
 MIGHTSILVFLILLIYLSLGRSLVPLVPLWNTANGAGIRGIPGPHOAIVASRDGH
 SRGALVVAAILGITAGTGAAPFNVGVWPLALSSALLALLLCLKLIGSTA"
 complement (2924..5107)
 /gene="Sma2379"
 complement (2924..5107)
 /gene="Sma2379"
 /EC_number="1.11.1.6"
 /functions="Cell Processes; Protection responses;
 detoxification"
 /notes="glimmer prediction; highly similar to
 catalase/peroxidase from Streptomyces reticuli, CAA74698"
 /codon_start=1
 /transl_table=11
 /product="catalase/peroxidase"
 /protein_id="AAK65944.1"
 /db_xref="GI:14524460"
 /translation="MQKSDSAGKCPVAHTAPGRSNRDNWPDQLDVQVLRHHSGLSD
 PLGNTFYAEFFKLLDLDALRDLRALMTDSQWPFADFGHGLGFIRMAWHSAGTYR
 ITDGRGAGQQRFPAPLNSPDNANLDKARLLWPIKQYGNRISSWADLLITGNVA
 LESMGKFTFGAGRVDMVPELFWGPBGTLGDBRYSGERQLSPALAAVQMLIYV
 NPEPGNPDVAAARDIRTFARMAANDEETVALLIAGGTFGKTHGAGDPGFIGADP
 EGGAIEDOGLGWKSTGTVGKDAITGGPEVTWSOTPTKSNHFFENLHHEWELTKS
 PAGAYQWAKNAEATIPDAYDPSRKHVPTFLTDLSLRDPAYEKLSRSLFENPDFA
 DAFARWFKLTHRDMPKSTVPEPAEDLIWQDVPVPAVDHRLVDETIDIAKAKII
 ASGLSVOELVSTASASTFRGSDKRGANGARIRLAPQKDMVNPAPQALRVSLVE
 GIORDFAQTDGKISLADLIVLAGAAVEAKAKAGHDITVPPTFGMDASEAQT
 AASFAALEPRADGRFNVYSTRQCFMKPEALVDAQLLTLTAPETVLTGGLRLVKA
 GEPKHGVTSRPEALTNDFVNLDMGTOWSPIEGEGVVEGREDRTGAARTGTTRVD
 LIIGSRHQLEAPAEVYAQSDAREKFKVDVFAVWTKVYNADRPDLV"
 5640..6113 /gene="Sma2381"
 5640..6113 /gene="Sma2381"
 /functions="Miscellaneous; Unknown"
 /notes="glimmer prediction"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK65945.1"
 /db_xref="GI:14524461"
 /translation="MLLKAMKIGRQSLYDTFGKWKLYCIAYERSASSETSALHALLR
 GKPRALEGIRLVNVRVVDNANOACLVNSICEFGQSRPDLAALHHAADRLKXNVVER
 IREAOAGDLAVSLPAEAVADFLVANIAGIRIARGGARREHLQSLRLALRAVT"
 6167..7135 /gene="Sma2383"
 6167..7135 /gene="Sma2383"
 /functions="Small molecule metabolism"
 /notes="glimmer prediction; similar to TED2, a putative
 quinone oxidoreductase, from Zinnia elegans, BAA06460, and
 from cowpea, T11672; also similar to putative
 oxidoreductases from Mycobacterium tuberculosis,
 Agrobacterium tumefaciens, etc. Zinc-binding family"
 /codon_start=1
 /transl_table=11
 /product="probable oxidoreductase"
 /protein_id="AAK65946.1"
 /db_xref="GI:14524462"
 /translation="MKAVMKVEGTDVMEFVDPRPFPVAPRHVVVEVAAAGVNFMDI
 GVFGQAWTDIPNPKVLVEGAGRVLAGDGTGEFAGVGRVAVVYAPGSAQRCSIPA
 ASLVKIPDVTDDRTAAATMQGLTASHFATDFYPQVGDIALVHAAGVGGLLTGII
 RLRGGRVIGRVSSSDKVAIRAKAGAEHVITDTRFADBLRLTGGGVNVYDGGSP
 KTFKGSIEALRRSGTFCWYGLGPGPLEIMNLPKSIKIGYATFMHTRELLDR
 TKOLFWDIEDGSITVTIGTYRLADAAGAHAAMASRATTGKLLIP"
 complement (7157..8946)
 /gene="Sma2385"

```

CDS
    complement(7357..8946)
    /gene="Sma2385"
    /function="Cell processes; Transport of small molecules"
    /note="glimmer prediction; similar to probable ABC
    transporter from Streptomyces coelicolor, T35245, and to
    VarM (probable transporter for virginiamycin) from
    Streptomyces virginiae, BAA96297. ABC transporter family"
    /codon_start=1
    /transl_table=11
    /product="probable ABC transporter, ATP-binding protein"
    /protein_id="AAK65947.1"
    /db_xref="GI:14524463"
    /translation="MPSITLSALSWKPDGEHVSFSLDLAFGPBRTGLVGRNGIGKTS
    VLNIAETLRSSGTVAIQGRVALARQILRAGADETADVFCAQCAVAVLRAEKQDA
    SVEIETADWTEVERISALARGLEARDILLNOLSGGQTRAVLAAAFSEPDFLL
    IDEPTNNIDRGRRAVIGLLSGWRSAGVIVSHRELLEEMDAIIELSLGTRKGGGM
    SAYQAARAELEAAQOQSLTARKTADVDKRALAEERLDKRDASGTRKAAKGMPRI
    LVGRKKSASRSGVELAEERRAGALDAVTAAKARLEVQPFSLPRLTELPAQRGT
    VLAPGVAGVDPARPIIRDLFSVLGPRVSVTPGNSGKTSILKVTGLPFPKGT
    VSNVVPFTLLDQSVLSLGERGTILENFKRLNPGASDNACRAALASFRADAALQVSE
    ALSGGQVLRAGLACALGSDPSPSLIIDEPTNHLIDISIEAVEAGLLSYDGLVTVSH
    DETFLANIGIGTRVELSTSRG"
    gene
    9057..9713
    /gene="Sma2387"
    CDS
    9057..9713
    /gene="Sma2387"
    /function="Miscellaneous; Unknown"
    /note="glimmer prediction; putative bacterial regulatory
    protein, tetr family"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAK65948.1"
    /db_xref="GI:14524464"
    /translation="MIRARRAGRPTRREAEALTRLLDSARSTFARKGIANGSMEE
    IAAELGSKHLYRYPNRQALLRAVVERDLVRFKTLAEAGGGEAPLAALRDMAR
    YFRFGTDYSAFLSVTAEEVFLPLERLAWSAALEPLVQAIISAQAAGLVVFG
    STIETCHVLIDLEGANVRVLCISSESPDASERLRFESRWAVQTAMMPZPNRPLGV
    "
    gene
    complement(9922..10455)
    /gene="Sma2389"
    CDS
    complement(9922..10455)
    /gene="Sma2389"
    /function="Cell processes; Adaptation"
    /note="glimmer prediction; very similar to Ohr, a putative
    stress-induced protein, from Bradyrhizobium japonicum,
    Query Match 87.0%; Score 17.4; DB 1; Length 10578;
    Best Local Similarity 94.7%; Pred. No. 1e+03;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGCA 19
    |||||
    Db 1017 CGAGGCGCCCGCCGCA 999

RESULT 9
AE007285
LOCUS
DEFINITION
    Sinorhizobium meliloti plasmid pSymA section 91 of 121 of the
    complete plasmid sequence.
ACCESSION
    AE007285 AB006469
VERSION
    AE007285.1 GI:14524132
KEYWORDS
    Sinorhizobium meliloti (Rhizobium meliloti)
SOURCE
    Sinorhizobium meliloti
    Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
    Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 10931)
REFERENCE
    Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
    Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
    Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
    Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
    Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
    Direct Submission
    Submitted (29-MAR-2001) Biological Sciences, Stanford University,
    371 Serra Mall, Stanford, CA 94305, USA
    Location/Qualifiers
        1..10931
        /organism="Sinorhizobium meliloti"
        /mol_type="genomic DNA"
        /strain="1021"
        /db_xref="taxon:382"
        /plasmid="pSymA"
        complement(218..568)
        /gene="Sma1791"
        complement(218..568)
        /gene="Sma1791"
        /function="Miscellaneous; Unknown"
        /note="glimmer prediction"
        /codon_start=1
        /transl_table=11
        /product="Hypothetical protein"
        /protein_id="AAK65646.1"
        /db_xref="GI:14524133"
        /translation="MDVGLPCITSSPPGCGSRAAATAGRVPRRLSTRVAARCCRP
        I
        NTRPVPVAPRASV"
        962..1384
        /gene="Sma1793"
        962..1384
        /gene="Sma1793"
        /function="Miscellaneous; Unknown"
        /note="glimmer prediction"
        /codon_start=1
        /transl_table=11
        /product="Hypothetical protein"
        /protein_id="AAK65647.1"
        /db_xref="GI:14524134"
        /translation="MEAVNRSTLALGLTMAATPLIAWVTPAAATYGPDEGEIGPG
        VRVVALGERASVTPAYKAVKLRDVVIOAGAKTFDNVATNDMLCHMTGELSYYQNEKK
        FTVKGDVWTCAKADTTEGTQNTSNTSVAINRIIDLMTS"
        2609..2935
        /gene="Sma1797"
        2609..2935
        /gene="Sma1797"
        /function="Miscellaneous; Unknown"
        /note="glimmer prediction"
        /codon_start=1
        /transl_table=11
        /product="Hypothetical protein"
        /protein_id="AAK65648.1"
        /db_xref="GI:14524135"
        /translation="MSPKSRAPQAYRPAQGGIMADAQMKKLIEDCERDIAACLHPE
        SGIPETHLLQQLSLRLDGGQAKALGDWQGRWHDPGDDMSPSPLWWTQPEFFGT
        IWRLLKI"
        3551..5452
        /gene="kup2"
        /note="synonym: Sma1798"
        3551..5452
        /gene="kup2"
        /function="Cell processes; Transport of small molecules;
        cations"
        /note="Similar to LOCUS KUP_ECOLI a 622 aa which encodes
        the KUP SYSTEM POTASSIUM UPTAKE PROTEIN. ACCESSION
        Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
        Nucleotide sequence and predicted functions of the entire
        Sinorhizobium meliloti pSymA megaplasmid
        Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
        21396509
        PUBMED
        11481432
        2 (bases 1 to 10931)
        Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
        Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
        Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
        Kalman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
        Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
        Direct Submission
        Submitted (29-MAR-2001) Biological Sciences, Stanford University,
        371 Serra Mall, Stanford, CA 94305, USA
        Location/Qualifiers
            1..10931
            /organism="Sinorhizobium meliloti"
            /mol_type="genomic DNA"
            /strain="1021"
            /db_xref="taxon:382"
            /plasmid="pSymA"
            complement(218..568)
            /gene="Sma1791"
            complement(218..568)
            /gene="Sma1791"
            /function="Miscellaneous; Unknown"
            /note="glimmer prediction"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein"
            /protein_id="AAK65646.1"
            /db_xref="GI:14524133"
            /translation="MDVGLPCITSSPPGCGSRAAATAGRVPRRLSTRVAARCCRP
            I
            NTRPVPVAPRASV"
            962..1384
            /gene="Sma1793"
            962..1384
            /gene="Sma1793"
            /function="Miscellaneous; Unknown"
            /note="glimmer prediction"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein"
            /protein_id="AAK65647.1"
            /db_xref="GI:14524134"
            /translation="MEAVNRSTLALGLTMAATPLIAWVTPAAATYGPDEGEIGPG
            VRVVALGERASVTPAYKAVKLRDVVIOAGAKTFDNVATNDMLCHMTGELSYYQNEKK
            FTVKGDVWTCAKADTTEGTQNTSNTSVAINRIIDLMTS"
            2609..2935
            /gene="Sma1797"
            2609..2935
            /gene="Sma1797"
            /function="Miscellaneous; Unknown"
            /note="glimmer prediction"
            /codon_start=1
            /transl_table=11
            /product="Hypothetical protein"
            /protein_id="AAK65648.1"
            /db_xref="GI:14524135"
            /translation="MSPKSRAPQAYRPAQGGIMADAQMKKLIEDCERDIAACLHPE
            SGIPETHLLQQLSLRLDGGQAKALGDWQGRWHDPGDDMSPSPLWWTQPEFFGT
            IWRLLKI"
            3551..5452
            /gene="kup2"
            /note="synonym: Sma1798"
            3551..5452
            /gene="kup2"
            /function="Cell processes; Transport of small molecules;
            cations"
            /note="Similar to LOCUS KUP_ECOLI a 622 aa which encodes
            the KUP SYSTEM POTASSIUM UPTAKE PROTEIN. ACCESSION

```

```

P30016. "
/codon_start=1
/transl_table=11
/product="Kup2 Potassium uptake protein"
/protein_id="AAK65649.1"
/db_xref="GI:14524136"
/translation="WADSLDHAPQAQNNLPQFLATIGAVGVYGDIGTSPLYAPREA
LPPGPGVGRDEVIGLSLVLTATVITIKVLFLLRADNDGSGGTLSTLALLKK
GTYKPYLMPFAGVGLGALPIGAMITPALSIVLSAVEGLKLVAPALHDVLPISVIL
LUFVOSTGATVSVFPGITLVLWMAAGVAHIGDAILLSFNLNIGFLWNA
GLIGFVLGAITLVTEALYADLGHFRHSIQAWFAVFPALALNYLGGALVLS
HPDAINPPFLPFWALLPVLITATAATIASQVITCAFSLIOALHGLFLPRFEI
CYTSETQIQVLPVITLVLALMLFSGSEALAPAYGVISITGAMVITDITLAFE
FYRWGMPALTAIALLPLENLEVLGALNKVHGQVYPIIAGTLITVAVLHNKIKH
GVSVREKTARODIPLSQMAIVERKSHAPVQVGTAFELTATPTTTPAVLHNKIKH
NHLVHONVIMIKTARVYVPEKORYTTIKLSDRFSLELRFEGMDQNVSRALARC
REGKPFELMSTSYLGRKRLIADFGSLPQWQDKLFAMADSAIDPTEYFHLPNRV
VBLGEQVIT"
/complement(6078..6583)
/standard_name="ISRM2011-2; ISRM11"
/notes="Sna3014; fragment of insertion sequence; predicted
by homology"
/evidence="not experimental"
/rpt_family="ISRM2011-2/ISRM11"
/genes="Sma1799"
/complement(6092..6565)
/genes="Sma1799"
/complement(6092..6565)
/genes="Sma1799"
/functions="elements of external origin; Transposon-related
functions"
/codon_start=1
/transl_table=11
/product="Trm2011-2b transposase"
/protein_id="AAK65650.1"
/db_xref="GI:14524137"
/translation="MAPLPGMAPRGLVGVAPFGHWTMTFVAALRADRVSAFFILD
GPNGRRPIYVQVVLVPLKAGDIVLDNLGSHGQEIARAIRKAGARLEPLKYSF
DLNPIEKFAKIKHMLREKQAQRDAIHDELRLHQVTPQECAYFKEAGYERA"
/complement(6661..7920)
/genes="Sma1801"
/complement(6661..7920)
/genes="Sma1801"
/functions="elements of external origin; Transposon-related
functions"
/notes="identical (99%) to reverse transcriptase/maturase
[Sinorhizobium meliloti]. ACCESSION CAA72334."
/codon_start=1
/transl_table=11
/product="Reverse transcriptase, maturase"
/protein_id="AAK65651.1"
/db_xref="GI:14524138"
/translation="MTSEFTDKPRIKRRVIEYKAVKANRGAAGVGGTLEIFEK
DLAANLYKWNRSXSGTYEPFPVRAVSPKKGAGSERVLGVTVDRIAQMVYKQMIKP
DLDSLPLPSYGYRPGKSDADAVGTQRCKYDQVLEDFDKGLFNPLHLLKAVR
KDVKNWALLYIERMLTAMENKNGEIVERSRGTPQGVVSPILANFLHYAFDLWMT
THPDLPCWYADGLVHCQSEQOALVELSRLAACLOLWHPKTKIVYCKDORRR
EAYPNVTFDLQVQPRVANTQWDERFCGTPAVSPITALKMEATIKSLNIPTPTP
GTLAEATKQNLPLRGWIAVYGRSRLSLTLDVYNQKLRAWIERKFKFQSHKTRA
SIFLRKLARENGLFVHWKAGTNTFT"
/complement(8466..9015)
/standard_name="ISRM2011-2/ISRM11"
/notes="Sna3015; fragment of insertion sequence; predicted
by homology"
/rpt_family="ISRM2011-2/ISRM11"
/complement(8915..8922)
/genes="Sma1803"
/complement(8915..8922)
/genes="Sma1803"
/functions="elements of external origin; Transposon-related
functions"
/codon_start=1
/transl_table=11
/product="Trm2011-2a transposase"

```

```

/protein_id="AAK65652.1"
/db_xref="GI:14524139"
/translation="MARPSNDLREVVDAVTGEGLSCRAAAKRFGIGISTADMVRR
FRETGAAPQMGCHAPKRLSGPHRAWLLCRCRERDFTLHGLVAELSERGLKVDYRAV
WTFVHEGLSYKKRWSWPANGSGTSPATYCHDG"
/genes="Sma1806"
/genes="Sma1806"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65653.1"
/db_xref="GI:14524140"
/translation="MHPALERMQRGKRNLTALAADPVNFTGSELGKFAATTEMQOCD
QRISQSEPTVAQRLLIFRAHLPTNRGATVASRLRRADRLSRTERRRTRWDA
RRGRASSAWSAAPRPAYRCFFIAVARSASGPCNMILLIS"
/complement(9668..10072)
/genes="Sma1808"
/complement(9668..10072)
/genes="Sma1808"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction. Overlaps in opposite direction
with Sma1807"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65654.1"
/db_xref="GI:14524141"
/translation="MEISFAPQSRRCAPARTPRPKTSGIRSWFRGVHKLVDKQTCL
GGQSLVGVQPAQIASVSDASAKATSGDRLRSSATASRRPTPISHRYAASTRYPA
RPSASCPWPCSTSLTRKEHSGSALAPIQGV"
/complement(10090..10797)
/genes="Sma1809"
/complement(10090..10797)
/genes="Sma1809"
/EC_number="1.1.1.10"

Query Match      87.0%   Score 17.4;   DB 1;   Length 10931;
Best Local Similarity 94.7%   Pred. No. 1e+03;
Matches 18;   Conservative 0;   Mismatches 1;   Indels 0;   Caps 0;

QY  1  CGAGGCGCCGAGCCGCA 19
DB  6537 CGAGGCGCCGAGCCGCA 6555

RESULT 10
LOCUS      BTNXGENE               12706 bp      mRNA      linear      MAM 03-MAR-1999
DEFINITION B.taurus mRNA for tenascin-X.
ACCESSION  Y11915
VERSION    Y11915.1  GI:2462978
KEYWORDS   flexilin; tenascin-X; TN-X gene.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE  1 (bases 1 to 12706)
            Eleftheriou, F., Exposito, J. Y., Garrone, R. and Lethias, C.
            Characterization of the bovine tenascin-X
            J. Biol. Chem. 272 (36), 22866-22874 (1997)
            MEDLINE 97426436
            PUBMED 9278449
REFERENCE  2 (bases 1 to 12706)
            Eleftheriou, F.
            Direct Submission
            Submitted (17-MAR-1997) F. Eleftheriou, Institut de Biologie et
            Chimie des Protéines, CNRS, UPR-412, 7 passage du Vercors, 69367
            Lyon Cedex 07, FRANCE

```


Chromosome 16 libraries constructed by Norman Doggett

(unpublished). VECTOR: sCos-1
 IMPORTANT: This sequence is not the entire insert of clone
 LA16-316G12 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone LA16-316G12 is at 1 in this sequence.
 The true left end of clone LA16-399E4 is at 37377 in this sequence.
 The true right end of clone LA16-358B7 is at 21289 in this
 sequence.

FEATURES

source

1. 37476
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosomes="16"
 /clone="LA16-316G12"
 /clone_lib="LA16"

misc_feature

987. 2805

/notes="CpG island"

/evidence="not experimental"

join(1727. 1820,6611. 6751,6973. 7060,7235. 7315,
 7567. 7674,8993. 9052,9200. 9266,9365. 9560,9833. 9937,
 10057. 10134,10252. 10354,10598. 10705,10790. 10885,
 10984. 11087,11464. 11584,12093. 12217,12298. 12388,
 12540. 12596,12672. 12769,12869. 12920,13040. 13170,
 13320. 13443,13818. 13961,14057. 14149,14241. 14432,
 14513. 14581,14676. 14755,14918. 15026,15369. 15440,
 15616. 15696,15772. 15892,15965. 16088,16167. 16372,
 16450. 17510)
 /genes="c316G12.1"

mRNA

join(1727. 1820,6611. 6751,6973. 7060,7235. 7315,
 7567. 7674,8993. 9052,9200. 9266,9365. 9560,9833. 9937,
 10057. 10134,10252. 10354,10598. 10705,10790. 10885,
 10984. 11087,11464. 11584,12093. 12217,12298. 12388,
 12540. 12596,12672. 12769,12869. 12920,13040. 13170,
 13320. 13443,13818. 13961,14057. 14149,14241. 14432,
 14513. 14581,14676. 14755,14918. 15026,15369. 15440,
 15616. 15696,15772. 15892,15965. 16088,16167. 16372,
 16450. 17510)
 /genes="c316G12.1"
 /product="c316G12.1 (KIAA0734 (C2 domain protein))"
 /notes="match: cDNAs: Em:AB018277 Em:AF038191
 match: ESTs: Em:T77480 Em:H18551 Em:F13300 Em:R61025
 Em:R24790 Em:R85179 Em:A1217078 Em:AA078749 Em:AI696810
 Em:AI086907 Em:H15802 Em:AA293357 Em:H46334 Em:D81844
 Em:AI301780 Em:N45522 Em:AA782001 Em:R60906 Em:AA716721
 Em:AI613004 Em:AA716524 Em:C00805 Em:AI111375 Em:AA481839
 Em:W01878 Em:AI146703 Em:AA627545 Em:AA079823 Em:A1240952
 Em:AA258690 Em:AA482458 Em:AA682952 Em:AI638343
 Em:AI693236 Em:AI537399 Em:H18444 Em:AI358964 Em:AI564977
 Em:AA302073"
 /evidence="not experimental"

CDS

join(6621. 6751,6973. 7060,7235. 7315,7567. 7674,
 8993. 9052,9200. 9266,9365. 9560,9833. 9937,10057. 10134,
 10252. 10354,10598. 10705,10790. 10885,10984. 11087,
 11464. 11584,12093. 12217,12298. 12388,12540. 12596,
 12672. 12769,12869. 12920,13040. 13170,13320. 13443,
 13818. 13961,14057. 14149,14241. 14432,14513. 14581,
 14676. 14755,14918. 15026,15369. 15440,15616. 15696,
 15772. 15892,15965. 16088,16167. 16372,16450. 16553)
 /genes="c316G12.1"
 /codon_start=1
 /evidence="not experimental"

/product="c316G12.1 (KIAA0734 (C2 domain protein))"
 /protein_id="CB856182.1"
 /db_xref="GI:5912546"
 /db_xref="SPRMBL:094839"

/translation="MTLLDIKSSVLRVQVQVCPSPRRTEQDPSGASADPQEPATGAW
 KPGDGVFFAHMELMLKKGQGFPCLEVRSGSPAPPEVDFSLGRALAAPEVE
 MYEALYTVLRAGTMDQVDDEALSYLQVFGTSLEHTAIEVRKAKATPY
 ALKVSVWRANLLAKDPNGSDPFCMLGILPASDATREPAQKQRFKRGSKRGGP
 LPAKCTOVTEVKSSTLNPWKHFLEFIEDVSTDLHLIDHDDVDLSLVACRKLNE
 VIGLKGMYRFXQIVKSARANGTAGTBTEDTDLGLCLNI PVREVPVAGVDRWPKLEP

misc_feature

7785. 8651

/genes="c316G12.1"

/notes="Tandem repeat. Forced join. Region of tandem

repeat, typical sequence of each element:

GGTGACACACACACACACAGGTGCA Restriction digest data
 (BAMHI) suggest approximately 100 base pairs are missing
 from this assembly."

7788. 8686

/notes="29 copies 31 mer 65% conserved"

7817. 8048

/notes="8 copies 29 mer 72% conserved"

7819. 7953

/notes="5 copies 27 mer 70% conserved"

7937. 8299

/notes="11 copies 33 mer 66% conserved"

7968. 8270

/notes="3 copies 101 mer 72% conserved"

8224. 8397

/notes="6 copies 29 mer 70% conserved"

8417. 8474

/notes="2 copies 29 mer 98% conserved"

8450. 8690

/notes="7 copies 33 mer 68% conserved"

8478. 8635

/notes="2 copies 29 mer 93% conserved"

8538. 8673

/notes="4 copies 29 mer 75% conserved"

8592. 8617

/notes="13 copies 2 mer ca 92% conserved"

8596. 8676

/notes="3 copies 27 mer 84% conserved"

8657. 8700

/notes="22 copies 2 mer ac 75% conserved"

11087. 11288

/notes="101 copies 2 mer 99 56% conserved"

11215. 11313

/notes="3 copies 33 mer 77% conserved"

14773. 14974

/notes="2 copies 101 mer 93% conserved"

15751. 16315

/genes="c316G12.1"

/notes="CpG island"

/evidence="not experimental"

complement(17101. 17510)

/notes="match: STS: Em:R60906"

complement(join(17311. 17679,17960. 18023,18129. 18305,

18878. 19071,19450. 19669,19764. 19875))

/genes="c316G12.2"

complement(join(17311. 17679,17960. 18023,18129. 18305,

18878. 19071,19450. 19669,19764. 19835))

/genes="c316G12.3"

/product="c316G12.2 (novel protein similar to predicted

yeast worm and archae-bacterial proteins)"

/notes="match: cDNAs: Em:AI109401

match: ESTs: Em:AI435098 Em:AI418486 Em:AI660316

Em:AI143673 Em:AI435931 Em:AI693236 Em:AI091298

Em:AI675241 Em:AA215709 Em:AI524203 Em:AI684711

Em:AI571317 Em:AI094880 Em:AA317548 Em:AI341540 Em:H98014

```

Em:AA716721 Em:AI659324 Em:AI224948 Em:AA410348
Em:AI086907 Em:AI304322 Em:AI146703 Em:W01878 Em:AI111375
Em:AI262869 Em:C00805 Em:H46334 Em:D81844 Em:AI696810
Em:AI716524 Em:AI217078 Em:AA079823 Em:AA627545
Em:AA483839 Em:AI560264 Em:AA258690 Em:AI240952
Em:AA482458 Em:HI5902 Em:AI302073 Em:AI423636 Em:R45522
Em:AA293357 Em:AI638343 Em:AI301780 Em:R60906 Em:AI672681
Em:AI613004 Em:AA782001 Em:F31747 Em:AI537389 Em:HI8444
Em:AI358964 Em:AI564977
/evidence=not_experimental

Query Match      87.0%; Score 17.4; DB 9; Length 37476;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGGCCCGCCGCCACA 19
Db 6760 CGAGGGCCCGCCGCCACA 6778
|||||
|||||

RESULT 12
AF003626/c
LOCUS
DEFINITION
Homo sapiens chromosome X clone ICRFXc104-M0525, LLOXNC01-C1233,
ICRFXC104-Cl284, Qc-3C1, LLOXNC01-B1439, LLOXNC01-220B3,
LLOXNC01-57B6, Qc-12C11 map q28, complete sequence.
ACCESSION
AF003626
VERSION
AF003626.2 GI:27476107
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 86834)
Bleichschmidt, K., Rosenthal, A. and Platzer, M.
Chromosome X genomic sequence
Unpublished
2 (bases 1 to 86834)
Bleichschmidt, K., Drescher, B., Nordsiek, G., Schattevoy, R., Knop, A.,
Rosenthal, A., Kioschis, P. and Poustka, A.
Direct Submission
Submitted (09-MAY-1997) Genome Analysis, Institute for Molecular
Biotechnology, Beutenbergstrasse 11, Jena D 07745, Germany
Platzer, M.
3 (bases 1 to 86834)
Direct Submission
Submitted (03-JAN-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
On Jan 3, 2003 this sequence version replaced gi:2121302.
-----
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
-----
Project Information
Center project name: x99+
Center clone name: ICRFXc104-M0525 to Qc-12C11
-----
Summary Statistics
Sequencing vector: M13, pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 86833 bases at least Q40
Consensus quality: 86833 bases at least Q30
Consensus quality: 86834 bases at least Q20
Quality coverage: 24.18%
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.

```

----- Neighboring sequence information: This entry is overlapped by RP11-316K19, RP4-687A5, LLOXNC01-223D9 and RP4-741010. -----

Sequence Quality Assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file. -----

FEATURES Source

```

Location/Qualifiers
1..86834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="q28"
/clone="ICRFXC104-M0525, LLOXNC01-C1233, ICRFXc104-Cl284,
Qc-3C1, LLOXNC01-B1439, LLOXNC01-220B3, LLOXNC01-57B6,
Qc-12C11"
1..11008
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="LLOXNC01-223D"
/clone_lib="Lawrence Livermore human X cosmid library"
/notes="overlapping clone"
1..22000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-316K1"
/clone_lib="RPCI human BAC 11"
/notes="overlapping clone"
1..26252
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP4-687A"
/clone_lib="RPCI human PAC 4"
/notes="overlapping clone"
1..45419
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="ICRFXC104-M0525"
/clone_lib="ICRF human X specific cosmid library no.104"
2178..45419
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="LLOXNC01-C1233"
/clone_lib="Lawrence Livermore human X cosmid library"
10530..53681
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="ICRFXC104-Cl284"
/clone_lib="ICRF human X specific cosmid library no.104"
11279..47432
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Qc-3C1"
/clone_lib="QIZ-derived Xq27.3-Xqter cosmid library"
14100..59047
/organism="Homo sapiens"
/mol_type="genomic DNA"

```

```

source
/db xref="taxon:9606"
/clone="LLOXNC01-B1439"
/clone_lib="Lawrence Livermore human X cosmid library"
43701..80327
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="LLOXNC01-22033"
/clone_lib="Lawrence Livermore human X cosmid library"
49823..86834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="Qc-12C11"
/clone_lib="Q12-derived Xq27.3-Xqter cosmid library"
52548..86834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="RP4-74101"
/clone_lib="RPCI human PAC 4"
/notes="overlapping clone"
74537..83583
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db xref="taxon:9606"
/clone="LLOXNC01-57B6"
/clone_lib="Lawrence Livermore human X cosmid library"
1029..11048
/notes="deleted in clone: RP4-687A5, deleted in clone:
LLOXNC01-223D9, deleted in clone: ICRFXC104-M0525"
/replace="G"
variation
1537
/notes="A substituted in clone: RP11-316K19"
/replace="G"
3051
/notes="T substituted in clone: RP11-316K19"
/replace="G"
3156
/notes="C substituted in clone: RP11-316K19"
/replace="A"
4270..4273
/notes="deleted in clone: RP4-687A5, deleted in clone:
LLOXNC01-223D9, deleted in clone: LLOXNC01-C1233"
/replace="G"
4299..4306
/notes="deleted in clone: ICRFXC104-M0525"
/replace="G"
4302..4303
/notes="deleted in clone: RP4-687A5, deleted in clone:
LLOXNC01-C1233, deleted in clone: LLOXNC01-223D9"
/replace="G"
4336..4361
/notes="deleted in clone: RP4-687A5"
/replace="G"
4350..4361
/notes="deleted in clone: RP11-316K19"
/replace="G"
4378
/notes="deleted in clone: LLOXNC01-223D9, deleted in
clone: RP4-687A5"
/replace="G"
9658
/notes="T substituted in clone: LLOXNC01-C1233"
/replace="C"
11514
/notes="deleted in clone: LLOXNC01-C1233, deleted in
clone: ICRFXC104-C1284, deleted in clone:
ICRFXC104-M0525"
/replace="G"
12953
/notes="A substituted in clone: RP11-316K19"
/replace="G"

```

```

variation
13065
/notes="G substituted in clone: RP11-316K19"
/replace="T"

Query Match      87.0%; Score 17.4; DB 9; Length 86834;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGAGCCGCCACA 19
Db 16859 CGAGGTGCCCGAGCCGCCACA 16841

RESULT 13
AL354677/c
LOCUS Homo sapiens chromosome X clone RP11-479P15, 27 unordered pieces.
DEFINITION AL354677
ACCESSION AL354677.3 GI:9931730
VERSION HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Melay, K.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 28, 2000 this sequence version replaced gi:9863680.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA479P15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 78464 bases at least Q40
Consensus quality: 85960 bases at least Q30
Consensus quality: 90075 bases at least Q20
Insert size: 95234; sum-of-contigs
Insert size: 197762; agarose-fp
Quality coverage: 1.99X in Q20 bases; sum-of-contigs Quality
Coverage: 1.47X in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3673: contig of 3673 bp in length
* 3674 3773: gap of 100 bp
* 3774 6066: contig of 2293 bp in length
* 6067 6166: gap of 100 bp
* 6167 9708: contig of 3540 bp in length
* 9707 9808: gap of 100 bp
* 9807 21235: contig of 11429 bp in length
* 21236 21335: gap of 100 bp
* 21336 24749: contig of 3414 bp in length
* 24750 24849: gap of 100 bp
* 24850 27355: contig of 2506 bp in length
* 27356 27455: gap of 100 bp
* 27456 31116: contig of 3661 bp in length
* 31117 32116: gap of 100 bp
* 31217 35127: contig of 3911 bp in length

```



```
* 35128 35227: gap of 100 bp
* 35228 38448: contig of 3221 bp in length
* 38449 38548: gap of 100 bp
* 38549 40769: contig of 2221 bp in length
* 40770 40869: gap of 100 bp
* 40870 45236: contig of 4367 bp in length
* 45237 45336: gap of 100 bp
* 45337 47792: contig of 2436 bp in length
* 47793 47892: gap of 100 bp
* 47893 53266: contig of 5374 bp in length
* 53267 53366: gap of 100 bp
* 53367 58271: contig of 4905 bp in length
* 58272 58371: gap of 100 bp
* 58372 63797: contig of 5426 bp in length
* 63798 63897: gap of 100 bp
* 63898 66275: contig of 2378 bp in length
* 66276 66375: gap of 100 bp
* 66376 68543: contig of 2168 bp in length
* 68544 70687: contig of 2044 bp in length
* 70688 73405: contig of 2618 bp in length
* 73406 73505: gap of 100 bp
* 73506 76943: contig of 3438 bp in length
* 76944 77043: gap of 100 bp
* 77044 79123: contig of 2080 bp in length
* 79124 79223: gap of 100 bp
* 79224 83079: contig of 3856 bp in length
* 83080 83179: gap of 100 bp
* 83180 85577: contig of 2298 bp in length
* 85578 85777: gap of 100 bp
* 85778 89299: contig of 3352 bp in length
* 89300 91739: contig of 100 bp
* 91740 91839: contig of 2710 bp in length
* 91840 94358: contig of 2419 bp in length
* 94359 94359: gap of 100 bp
* 94360 97834: contig of 3476 bp in length.
* 94359 97834: contig of 3476 bp in length.

FEATURES
      source
      1..97834
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="X"
      /clone="RP11-479P15"
      /clone_lib="RPC1-11.2"
      1..3673
      /note="assembly fragment:00141"
      fragment_chain:1
      3774..5066
      /note="assembly fragment:00636"
      fragment_chain:1
      6167..5706
      /note="assembly fragment:00363"
      fragment_chain:2
      5807..21235
      /note="assembly fragment:00564"
      fragment_chain:2
      21336..24749
      /note="assembly fragment:00449"
      fragment_chain:3
      24850..27355
      /note="assembly fragment:00808"
      fragment_chain:3
      27456..31116
      /note="assembly fragment:00076"
      31217..35127
      /note="assembly fragment:00114"
      35228..38448
      /note="assembly fragment:00143"
      38549..40769
      /note="assembly fragment:00173"
      40870..45236
```

```
misc_feature /note="assembly_fragment:00176"
45337..47792
misc_feature /note="assembly_fragment:00225"
47893..53266
misc_feature /note="assembly_fragment:00265"
53367..58271
misc_feature /note="assembly_fragment:00355"
58372..63797
misc_feature /note="assembly_fragment:00400"
63898..66275
misc_feature /note="assembly_fragment:00423"
66376..68543
misc_feature /note="assembly_fragment:00460"
68544..70687
misc_feature /note="assembly_fragment:00461"
70788..73405
misc_feature /note="assembly_fragment:00464"
73506..76943
misc_feature /note="assembly_fragment:00536"
77044..79123
misc_feature /note="assembly_fragment:00586"
79224..83079
misc_feature /note="assembly_fragment:00631"
83180..85477
misc_feature /note="assembly_fragment:00707"
85578..88929
misc_feature /note="assembly_fragment:00773"
89030..91739
misc_feature /note="assembly_fragment:00773"
91840..94258
misc_feature /note="assembly_fragment:00830"
94359..97834
misc_feature /note="assembly_fragment:00894"
97835..100000

ORIGIN
```

```
Query Match 87.0%; Score 17.4; DB 2; Length 97834;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 CGAGGCGCCGAGCCGCCACA 19
```

```
Db 57840 CGAGGTGCCGAGCCGCCACA 57822
```

```
RESULT 14
```

```
AC092269 LOCUS 105589 bp DNA linear HTG 04-SEP-2001
AC092269 DEFINITION Homo sapiens chromosome 5 clone CTB-78L15, WORKING DRAFT SEQUENCE.
AC092269 ACCESSION AC092269
AC092269.2 GI:15290438
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
```

```
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 105589)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 105589)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
```

```
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:14589458.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
```

```
-----
Project Information
Center Project Name: 94435
Center clone name: CIT978SKB_78L15
```


McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mikova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M.,
Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 28, 2000 this sequence version replaced gi:7009813.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Project name: L4974
Center project name: L4974
Center clone name: 462_J_14

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1071: contig of 1071 bp in length
* 1072 1171: gap of 100 bp
* 1172 2441: contig of 1270 bp in length
* 2442 2541: gap of 100 bp
* 2542 3920: contig of 1379 bp in length
* 3921 4020: gap of 100 bp
* 4021 5112: contig of 1092 bp in length
* 5113 5212: gap of 100 bp
* 5213 6856: contig of 1644 bp in length
* 6857 6957: gap of 100 bp
* 6958 8429: contig of 1373 bp in length
* 8430 8429: gap of 100 bp
* 8430 10443: contig of 2014 bp in length
* 10444 10543: gap of 100 bp
* 10544 11871: contig of 1328 bp in length
* 11872 11971: gap of 100 bp
* 11972 12442: contig of 271 bp in length
* 12443 12342: gap of 100 bp
* 12343 14098: contig of 1756 bp in length
* 14099 14198: gap of 100 bp
* 14199 15712: contig of 1514 bp in length
* 15713 15812: gap of 100 bp
* 15813 18266: contig of 2454 bp in length
* 18267 18366: gap of 100 bp
* 18367 20290: contig of 1524 bp in length
* 20291 20390: gap of 100 bp
* 20391 22569: contig of 2179 bp in length
* 22570 22669: gap of 100 bp
* 22670 25385: contig of 2716 bp in length
* 25386 25485: gap of 100 bp
* 25486 28984: contig of 3399 bp in length
* 28985 31777: contig of 2793 bp in length
* 31778 31877: gap of 100 bp
* 31878 34560: contig of 2683 bp in length
* 34561 34660: gap of 100 bp
* 34661 37141: contig of 2481 bp in length
* 37142 37241: gap of 100 bp
* 37242 40350: contig of 3109 bp in length

* 40351 40450: gap of 100 bp
* 40451 43420: contig of 2970 bp in length
* 43421 43520: gap of 100 bp
* 43521 46331: contig of 2811 bp in length
* 46332 46431: gap of 100 bp
* 46432 49168: contig of 2737 bp in length
* 49169 49268: gap of 100 bp
* 49269 52366: contig of 3098 bp in length
* 52367 52466: gap of 100 bp
* 52467 56270: contig of 3804 bp in length
* 56271 59259: contig of 2889 bp in length
* 59260 63099: contig of 3740 bp in length
* 63100 63199: gap of 100 bp
* 63200 67884: contig of 4685 bp in length
* 67885 71879: contig of 3895 bp in length
* 71880 71979: gap of 100 bp
* 71980 77510: contig of 5531 bp in length
* 77511 77610: gap of 100 bp
* 77611 81594: contig of 3984 bp in length
* 81595 87445: contig of 5751 bp in length
* 87446 87545: gap of 100 bp
* 87546 92288: contig of 4743 bp in length
* 92289 92388: gap of 100 bp
* 92389 97884: contig of 5495 bp in length
* 97885 97983: gap of 100 bp
* 97984 104615: contig of 6632 bp in length
* 104616 112913: contig of 8198 bp in length
* 112914 113013: gap of 100 bp
* 113014 121166: contig of 8153 bp in length
* 121167 121266: gap of 100 bp
* 121267 129653: contig of 8387 bp in length
* 129654 139142: contig of 9389 bp in length
* 139143 139242: gap of 100 bp
* 139243 148958: contig of 9716 bp in length
* 148959 149058: gap of 100 bp
* 149059 160677: contig of 11618 bp in length
* 160677 160777: gap of 100 bp
* 160777 175511: contig of 14735 bp in length.

FEATURES
source
1. .175511
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-462J14"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. .1071
/note="assembly_fragment"
misc_feature
1172. .2441
/note="assembly_fragment"
misc_feature
2542. .3920
/note="assembly_fragment"
misc_feature
4021. .5112
/note="assembly_fragment"
misc_feature
5213. .6856
/note="assembly_fragment"
misc_feature
6957. .8329
/note="assembly_fragment"
misc_feature
8430. .10443
/note="assembly_fragment"
misc_feature
10544. .11871
/note="assembly_fragment"
misc_feature
11972. .12242
/note="assembly_fragment"
clone_end:SP6
vector_side:right

Mon Jun 21 09:02:04 2004

```

misc_feature 12343..14098
/note="assembly_fragment"
misc_feature 14199..15712
/note="assembly_fragment"
misc_feature 15813..18266
/note="assembly_fragment"
misc_feature 18367..20290
/note="assembly_fragment"
misc_feature 20391..22569
/note="assembly_fragment"
misc_feature 22670..25385
/note="assembly_fragment"
misc_feature 25486..28894
/note="assembly_fragment"
misc_feature 28985..31777
/note="assembly_fragment"
misc_feature 31878..34560
/note="assembly_fragment"
misc_feature 34661..37141
/note="assembly_fragment"
misc_feature 37242..40350
/note="assembly_fragment"
misc_feature 40451..43420
/note="assembly_fragment"
misc_feature 43521..46331
/note="assembly_fragment"
misc_feature 46432..49168
/note="assembly_fragment"

Query Match 87.0%; Score 17.4; DB 2; Length 175511;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGCGCCGACGCCCCACAA 20
Db 168379 GAGCGCCGACGCCCCACAA 168397

RESULT 19
AC142187
LOCUS AC142187 181070 bp DNA linear HTG 24-MAR-2003
DEFINITION Rattus norvegicus clone CH230-368G11, WORKING DRAFT SEQUENCE, 28
unorderd pieces.
ACCESSION AC142187
VERSION AC142187.1 GI:29165337
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 181070)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blawalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Dalgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guvera, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

```

```

Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Loulsegad, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G.,
Olapunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E.,
Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E.,
Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A.,
Svatek, A., Tabor, P., Taylor, C., Taylor, R., Thomas, N., Thomas, S.,
Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,
Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc-bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KSKU
Center clone name: CH230-368G11
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167171 bases at least Q40
Consensus quality: 170222 bases at least Q30
Consensus quality: 172408 bases at least Q20
Estimated insert size: 172425; sum-of-contents estimation
Quality coverage: 4x in Q20 bases; sum-of-contents estimation

* NOTE: Estimated insert size may differ from sequence length
* (see this://www.hgsc-bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1309: contig of 1309 bp in length
* 1409: gap of unknown length
* 1410: contig of 1007 bp in length
* 2416: gap of unknown length
* 2417: gap of unknown length
* 2517: contig of 1836 bp in length
* 4352: contig of 1836 bp in length
* 4353: gap of unknown length
* 4452: gap of unknown length
* 5649: contig of 1197 bp in length
* 5650: gap of unknown length
* 5750: contig of 1527 bp in length
* 7276: contig of 1527 bp in length
* 7277: gap of unknown length

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT


```

* 67456 69673: contig of 2218 bp in length
* 69674 69773: gap of 100 bp
* 69774 72978: contig of 3205 bp in length
* 72979 73078: gap of 100 bp
* 73079 82532: contig of 9454 bp in length
* 82533 82632: gap of 100 bp
* 82633 95585: contig of 12953 bp in length
* 95586 95686: gap of 100 bp
* 95687 117061: contig of 21376 bp in length
* 117062 117161: gap of 100 bp
* 117162 149013: contig of 31852 bp in length
* 149014 149113: gap of 100 bp
* 149114 181372: contig of 32259 bp in length
* 181373 181472: gap of 100 bp
* 181473 182765: contig of 1283 bp in length.

```

FEATURES

```

source
1. .182765
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="8"
/map="8"
/clone="RP11-973F15"
/clone_lib="RP11-973F15 Human Male BAC"
1. .14137
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
14238..15333
/notes="assembly_fragment"
15434..16426
/notes="assembly_fragment"
16527..17835
/notes="assembly_fragment"
17936..67355
/notes="assembly_fragment"
67456..69673
/notes="assembly_fragment"
69774..72978
/notes="assembly_fragment"
73079..82532
/notes="assembly_fragment"
82633..95585
/notes="assembly_fragment"
95686..117061
/notes="assembly_fragment"
117162..149013
/notes="assembly_fragment"
149114..181372
/notes="assembly_fragment"
181473..182765
/notes="assembly_fragment"
clone_end:T7
vector_side:right

```

ORIGIN

```

Query Match 87.0%; Score 17.4; DB 2; Length 182765;
Best Local Similarity 94.7%; Pred.No.1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCCAACAA 20
|||||
Db 50578 GAGGCGCAGCCCAACAA 50596

RESULT 21
AC124791
LOCUS Homo sapiens chromosome 8 clone RP11-973F15 map 8, WORKING DRAFT
DEFINITION SQUENCE, 22 ordered pieces.
ACCESSION AC124791
VERSION AC124791.1 GI:21431212
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

```

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 8, clone RP11-973F15

Unpublished

2. (bases 1 to 186908)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Bonuskavkiy, L., Boukharov, B., Brown, A., Camarato, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lenockzy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27366

Center clone name: 973 F 15

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178955 bases at least Q40

Consensus quality: 182683 bases at least Q30

Consensus quality: 184121 bases at least Q20

Insert size: 124000; agarose-fp

Insert size: 184808; sum-of-contigs

Quality coverage: 5.8 in Q20 bases; agarose-fp

Quality coverage: 6.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 1372: contig of 1372 bp in length

* 1373 1472: gap of 100 bp

* 1473 3660: contig of 2188 bp in length

* 3661 3760: gap of 100 bp

* 3761 5017: contig of 1257 bp in length

* 5018 5117: gap of 100 bp

* 5118 7530: contig of 2413 bp in length


```

/genes="idnO1 OR SMD20692"
/EC number="1.1.1.69"
/function="Small Molecule Metabolism; Degradation; carbon
compounds"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative 5-Keto-D-gluconate 5-reductase protein"
/protein_id="CAC49799.1"
/db_xref="GI:15141287"
/db_xref="GOA:Q92TU6"
/db_xref="SPTREMBL:Q92TU6"
/translation="WMDGASNAAPRIRGIPALFDLSGRVALITSSGGILGTHMAHALC
EAGASVILNGRDEARLAGARALEBHGFTVGTSAFDVTKSAEIGRAVADILAERGRID
ILVNRAGIORTPLHEFPEDAPRRVIEINLTSAFLVGOAVQGMTEREGTILNCSV
QSELARPSIAPYAAASKGGLKMLTKGMALDWGRYGRVUNGARFCYFKTELSALVSDK
FSTWLEQRTPLGRMGDTGLAAAFLASAAASFVTHGILYVDGITSCL"
complement(1063..1361)
/genes="SMD20693"
complement(<1063..1361)
/genes="SMD20693"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49800.1"
/db_xref="GI:15141288"
/db_xref="GOA:Q92TU5"
/db_xref="SPTREMBL:Q92TU5"
/translation="MPDYAASKGGVQOLAKAFATNGQPCTFRSMELPPGYIATDSTEA
SNDPSRARSILARIIPRAAGACGRTLPARRSPWRPHPTTSRRSSPWTAAAGWDA"
1508..1684
/genes="SMD20694"
1508..1684
/genes="SMD20694"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49801.1"
/db_xref="GI:15141289"
/db_xref="GOA:Q92TU4"
/db_xref="SPTREMBL:Q92TU4"
/translation="WGAGDDPMAALDARVRVIERLRVADMSAVPDINAGNTNAPAT
MGNRCAGFILGTA"
complement(1765..2271)
/genes="SMD20695"
complement(1765..2271)
/genes="SMD20695"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49802.1"
/db_xref="GI:15141290"
/db_xref="GOA:Q92TU3"
/db_xref="SPTREMBL:Q92TU3"
/translation="MTLLPVLGGTGLVAVRLDQAVHAPTWDSEGEAVRVGRWNSKG
VRAVYCSLDPATIAIEVAVHGRFALDMVHTMTAAATADAGDVYVDPKSPVNPNA
RPGIPGAGQAGVGDLLRRHREVAIPSAVSPHSWNVLVFLGGAAPAAVAKPQEPFALD
TRLHPPTG"
complement(2268..2753)

```

```

/genes="SMB20696"
complement(2268..2753)
/genes="SMB20696"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49803.1"
/db_xref="GI:15141291"
/db_xref="GOA:Q92TU2"
/db_xref="SPTREMBL:Q92TU2"
/translation="MALAAKAPPAAGGELQKSIETLGGSRVLAHLHNLNDAHELL
LRGLPASALDLVGNLVFTGKTESLEKAVMSLRTWRRKETPNKPLSQEQSGRAWKF
AEILVKATDIFGSGAAEQWLERPAVLQRRRIDLGTTPPGVELVEDHLDRLEYGIY
A"
complement(2855..4549)
/genes="SMB20697"
complement(2855..4549)
/genes="SMB20697"
/function="MACROMOLECULE METABOLISM; Macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative peptidase, similar to arginine
utilization protein RcoB"
/protein_id="CAC49804.1"
/db_xref="GI:15141292"
/db_xref="GOA:Q92TU1"
/db_xref="SPTREMBL:Q92TU1"
/translation="MTNAQEBTVSRAGRSIDSRVQCFALRMTSPSETGTGCEASFA
DRLLGGLGELPYPREHPQDLGLLASHGEPLTRNVVALVRGTGKRTLVMAGHPTVSTD
NYHELKALCDSLALDKALIESLARTSEOEERALQDLASQDFLPGRLDLMKSLG
AVAIACLEQPAADTRQGNLMVLVATPEDEERSGNSRLDALPLGVDRFDIAGGIN
LDVTSDDGSGEGRVAVGTIGKLLPFALVIGCSHASYFFEGVSAQAVALKARLE
GNVADRDNDNDISPPPICLEAKLDGDEYVTFERFIANLWYHMTADALFERFR
EEVLTGANEAIEYFAAASAEYKLVGRSAGLPAPKPLLSFOELRAAARVFGHGFDA
FYAKEREFAQSDNPLVATROLTLMVGIARLSGPAIVIGFSGLHYPSPSHLSLEAGND
RSLHOALEKARAGLGNDRPALSILWKPHFYGISDMSPLGLAASGSQVVSNTPLSLVD
RPSENALRPTNVLGPWGREFHQKFERVHEPVAFRVLPELVSEIARTFLGDDRRHD"
complement(4701..6353)
/genes="SMB20698"
complement(4701..6353)
/genes="SMB20698"
/function="CELL PROCESSES; Transport of small molecules"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein, similar to E. coli
multidrug resistance protein EmrB"
/protein_id="CAC49805.1"
/db_xref="GI:15141293"
/db_xref="GOA:Q92TU0"
/db_xref="SPTREMBL:Q92TU0"
/translation="MTIATPSPRDISMAGLLVAGIVLAVITATEIASTVLSLGRGDI
IGDYATPDEFAWLDVGYTTFKLGFMASWLSRDPNVLVGSTLAMGACGTAAM
TARLDLLVALRMTIQGSGTLVVGQAIIFLTFPHSRQPLQAFAMGVSVAPIAP
ALQGLLDSQSWTWIFFSIPLAALAAAGLLADGVATRAORPFDWTFGLISGAL
FCFTYVFSQSGRDWEPRILWLAVIGTATLLAFLOQVLAQKQGLPDTFETEDF
CFATYVSVAGAAALFCSAFLIPSPAVSLAFTPDAGQLIPSGALFICALIATFLM
QLRVPFVAIVFPGILINWAMMUSGTSSESGADMKAAILLRNGLGLFLSLTILI
AFNSNRNLGILGINTGRQGLGIVSALITLIEHNVSAILLRNGLGLFLSLTILI
ADLTWITALLTAKGMDAAASGAASLGRVAVAGOSTVIAEDTAFNALGLLFLVTAAP
VLVGIKIGLARYAKAERAGRAAKPSGHHTLPKRVFAS"
complement(6350..7447)

```

```

CDS
/gene="SMB20699"
complement(6350..7447)
/gene="SMB20699"
/function="CELL PROCESSES; Transport of large molecules;
protein, peptide secretion"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative protein secretion protein, HlyD family,
similar to E. coli EmrK"

Query Match 87.0%; Score 17.4; DB 1; Length 189333;
Best Local Similarity 94.7%; Pred. NO. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCCCACA 19
Db 80571 CGAGGCGCCAGCCCCGCA 80589

RESULT 23
AF274573/c
LOCUS AF274573 189436 bp DNA linear PRI 01-JAN-2003
DEFINITION Homo sapiens chromosome X clone RP11-316K19, RP11-316M19, RP4-687A5
ACCESSION AF274573
VERSION AF274573.4 GI:27452915
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189436)
AUTHORS Galgoczy,P., Schilhabel,M., Rosenthal,A. and Platzer,M.
TITLE Chromosome X genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189436)
AUTHORS Platzer,M.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
REFERENCE 3 (bases 1 to 189436)
AUTHORS Platzer,M.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
REFERENCE 4 (bases 1 to 189436)
AUTHORS Platzer,M.
TITLE Direct Submission
JOURNAL Submitted (01-JAN-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
COMMENT On Jan 1, 2003 this sequence version replaced gi:21539160.
-----
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
-----
Project Information
Center project name: x131+
Center clone name: RP11-316K19, RP4-687A5
-----
Summary Statistics
Sequencing vector: M13, pUC18, 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 183203 bases at least Q40
Consensus quality: 189383 bases at least Q30
Consensus quality: 189436 bases at least Q20
Quality coverage: 18.73x
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate

```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Neighboring sequence information:

This entry is overlapped by QC-11G7, LLOXNC01-P0142, ICRFXC104-M0525, LLOXNC01-C1233, QC-3C1, LLOXNC01-B1439 and covers RP11-316K19, LLOXNC01-D0132, RP4-687A5, LLOXNC01-223D9 entirely. RP11-316M19 is identical to RP11-316K19.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

| FEATURES | Location/Qualifiers |
|----------|---|
| source | 1..189436 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="X" /map="q28" /clone="RP11-316K19, RP11-316M19, RP4-687A5" |
| source | 1..2112 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="QC-11G" /clone_lib="Q12-derived Xq27.3-Xqter cosmid library" /notes="overlapping clone" |
| source | 1..30055 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="LLOXNC01-P014" /clone_lib="Lawrence Livermore human X cosmid library" /notes="overlapping clone" |
| source | 1..185184 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="RP11-316K19" /clone_lib="RPCI human BAC 11" 13446..39763 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="LLOXNC01-D0132" /clone_lib="Lawrence Livermore human X cosmid library" |
| source | 85923..189436 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="RP4-687A5" /clone_lib="RPCI human PAC 4" 136581..174193 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="LLOXNC01-223D9" /clone_lib="Lawrence Livermore human X cosmid library" |
| source | 163186..189436 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="ICRFXC104-M052" |

```

/clone_lib="RCRF human X specific cosmid library no.104"
/note="overlapping clone"
165363. .189436
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="LLOXNC01-C123"
/notes="overlapping clone"
/note="overlapping clone"
174464. .189436
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Qc-3C"
/clone_lib="Q12-derived Xq27.3-Xqter cosmid library"
177285. .189436
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="LLOXNC01-B143"
/clone_lib="Lawrence Livermore human X cosmid library"
/note="overlapping clone"
674
/note="C substituted in clone: Qc-11G7"
/replace="T"
1677
/note="C substituted in clone: LLOXNC01-PO142"
/replace="T"
2522
/note="A substituted in clone: LLOXNC01-PO142"
/replace="G"
4257
/note="C substituted in clone: LLOXNC01-PO142"
/replace="T"
5340. .5341
/note="deleted in clone: LLOXNC01-PO142"
/replace="G"
5825
/note="G substituted in clone: LLOXNC01-PO142"
/replace="C"
6442
/note="deleted in clone: RP11-316K19"
/replace="A"
7042
/note="C substituted in clone: LLOXNC01-PO142"
/replace="T"
8639
/note="G substituted in clone: LLOXNC01-PO142"
/replace="A"
9064
/note="deleted in clone: RP11-316K19"
/replace="A"
9546. .9549
/note="deleted in clone: RP11-316K19"
/replace="A"
9595. .9598
/note="deleted in clone: RP11-316K19"
/replace="A"
10095. 10098
/note="deleted in clone: RP11-316K19"
/replace="A"
11036
/note="T substituted in clone: LLOXNC01-PO142"
/replace="C"
12039
/note="C substituted in clone: LLOXNC01-PO142"
/replace="A"
12684
/note="C substituted in clone: LLOXNC01-PO142"
/replace="G"
13364
/note="C substituted in clone: LLOXNC01-PO142"

```

```

/replace="T"
14099. .14115
/note="deleted in clone: RP11-316K19"
/replace="A"
15637
variation
variation
Query Match 87.0%; Score 17.4; DB 9; Length 189436;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 CGAGGCCGCCAGCCGCCACA 19
|||||
DB 180044 CGAGGTGCCAGCCGCCACA 180026
|||||
RESULT 24
AC100872/c
LOCUS AC100872 195563 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-973F15, complete sequence.
ACCESSION AC100872 GI:19033627
VERSION AC100872.2
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 195563)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE Homo sapiens chromosome 8, clone RP11-973F15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195563)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,I., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamarez,R., Landers,T., Lechoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norman,C.H., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollata,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thumann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 195563)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,I., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamarez,R.,
Landers,T., Lechoczy,J., Levine,R., Liu,G., Maclean,C.,
MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollata,V., Raymond,C.,
Ratta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

```

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2002 this sequence version replaced gi:17048242.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L21665

Center clone name: 973_F_15

FEATURES

----- Location/Qualifiers
1..195563
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-973P15"
/clone_lib="RC1-11 Human Male BAC"

Source

1..206
/rpt_family="MLT2B1"

repeat_region

complement(288..364)

repeat_region

/rpt_family="t2"

repeat_region

complement(1408..1548)

repeat_region

/rpt_family="MIR"

repeat_region

complement(2093..2268)

repeat_region

/rpt_family="MER53"

repeat_region

2413..2433

repeat_region

/rpt_family="t2"

repeat_region

/rpt_family="LIME4A"

repeat_region

complement(5183..5492)

repeat_region

/rpt_family="LIME4A"

repeat_region

5631..5809

repeat_region

/rpt_family="LTR16C"

repeat_region

complement(5945..6163)

repeat_region

/rpt_family="MIR"

repeat_region

6979..7329

repeat_region

/rpt_family="t2"

repeat_region

complement(7578..7894)

repeat_region

/rpt_family="MER33"

repeat_region

complement(8249..8381)

repeat_region

/rpt_family="Charlies"

repeat_region

8828..8929

repeat_region

/rpt_family="MIR"

repeat_region

10603..10697

repeat_region

/rpt_family="MER94"

repeat_region

10767..10927

repeat_region

/rpt_family="MIR"

repeat_region

11146..11215

repeat_region

/rpt_family="MLT1C"

repeat_region

11216..11562

repeat_region

/rpt_family="MLT1A1"

repeat_region

11563..11894

repeat_region

/rpt_family="MLT1C"

repeat_region

11991..12287

repeat_region

/rpt_family="AluX"

repeat_region

complement(13632..13768)

repeat_region

/rpt_family="MIR"

repeat_region

complement(13772..14294)

repeat_region

/rpt_family="MLT1H"

repeat_region

Query Match

87.0%; Score 17.4; DB 9; Length 195563;

repeat_region complement(14444..14662)
/rpt_family="MIR"
repeat_region complement(16196..16597)
/rpt_family="MLT1J"
repeat_region complement(16696..16781)
/rpt_family="MIR"
repeat_region 16782..16818
/rpt_family="(CA)n"
repeat_region complement(16819..16901)
/rpt_family="MIR"
repeat_region 17008..17031
/rpt_family="(TTTTTC)n"
repeat_region 17360..17578
/rpt_family="MIR"
repeat_region complement(17697..18012)
/rpt_family="AluX"
repeat_region 18426..18563
/rpt_family="MIR3"
repeat_region complement(19638..19736)
/rpt_family="MER5B"
repeat_region complement(20522..20968)
/rpt_family="MLT2A2"
repeat_region complement(21378..21431)
/rpt_family="MER112"
repeat_region 22096..22199
/rpt_family="MIR"
repeat_region 22452..23470
/rpt_family="LIMD2"
repeat_region 23632..23990
/rpt_family="MER33"
repeat_region complement(24349..24616)
/rpt_family="AluX"
repeat_region 24846..24869
/rpt_family="(TC)n"
repeat_region 24869..24901
/rpt_family="(TG)n"
repeat_region complement(25740..25804)
/rpt_family="AluY"
repeat_region 25823..26231
/rpt_family="MER6A"
repeat_region complement(26232..26389)
/rpt_family="AluJo"
repeat_region 26390..26421
/rpt_family="(TA)n"
repeat_region complement(26422..26567)
/rpt_family="AluJo"
repeat_region 26568..26786
/rpt_family="MER6A"
repeat_region 26878..26959
/rpt_family="MIR"
repeat_region complement(27427..27561)
/rpt_family="MLT1I"
repeat_region complement(27563..27793)
/rpt_family="THE1A"
repeat_region complement(27794..28080)
/rpt_family="THE1A-int"
repeat_region complement(28081..28427)
/rpt_family="THE1A"
repeat_region complement(28566..28667)
/rpt_family="MLT1I"
repeat_region complement(28883..29019)
/rpt_family="L2"
repeat_region complement(29054..29261)
/rpt_family="MIR"
repeat_region 29340..29493
/rpt_family="MER117"
repeat_region 30200..30351
/rpt_family="MER117"
repeat_region complement(30618..30744)
/rpt_family="L2"
repeat_region complement(30745..31330)

Quality coverage: 4.7 in Q20 bases; agarose-fp
 Quality coverage: 4.9 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1
* 1811 1910: contig of 1810 bp in length
* 1911 1910: gap of 100 bp
* 5011 5018: contig of 3108 bp in length
* 5119 5118: gap of 100 bp
* 7959 7959: contig of 2841 bp in length
* 7960 8059: gap of 100 bp
* 8060 11695: contig of 3636 bp in length
* 11696 11795: gap of 100 bp
* 11796 14271: contig of 2476 bp in length
* 14272 14371: gap of 100 bp
* 14372 22386: contig of 8015 bp in length
* 22387 22486: gap of 100 bp
* 22487 29746: contig of 7260 bp in length
* 29747 29846: gap of 100 bp
* 29847 37497: contig of 7651 bp in length
* 37498 37597: gap of 100 bp
* 37598 43589: contig of 5992 bp in length
* 43590 43689: gap of 100 bp
* 43690 50270: contig of 6581 bp in length
* 50271 50370: gap of 100 bp
* 50371 57887: contig of 7517 bp in length
* 57887 57987: gap of 100 bp
* 57988 65412: contig of 7425 bp in length
* 65413 65512: gap of 100 bp
* 65513 75598: contig of 10186 bp in length
* 75599 75798: gap of 100 bp
* 75799 85662: contig of 9864 bp in length
* 85663 85762: gap of 100 bp
* 85763 97087: contig of 11325 bp in length
* 97088 97188: gap of 100 bp
* 97188 113019: contig of 15832 bp in length
* 113020 130955: contig of 17836 bp in length
* 130956 131055: gap of 100 bp
* 131056 148692: contig of 17637 bp in length
* 148693 148793: gap of 100 bp
* 148793 165067: contig of 16275 bp in length
* 165068 165167: gap of 100 bp
* 165168 185335: contig of 20168 bp in length
* 185336 185435: gap of 100 bp
* 185436 202437: contig of 17002 bp in length.
  
```

FEATURES

source

```

1..202437
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-420J13"
/clone_lib="RPC1-11 Human Male BAC"
1..1810
/note="assembly_fragment"
1911..5018
/note="assembly_fragment"
5119..7959
/note="assembly_fragment"
8060..11695
/note="assembly_fragment"
11796..14271
/note="assembly_fragment"
14372..22386
/note="assembly_fragment"
  
```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature 22487..29746
/note="assembly_fragment"
misc_feature 29847..37497
/note="assembly_fragment"
misc_feature 37598..43589
/note="assembly_fragment"
clone_end:17
vector_side:left"
43690..50270
/note="assembly_fragment"
50371..57887
/note="assembly_fragment"
57988..65412
/note="assembly_fragment"
65513..75598
/note="assembly_fragment"
75799..85662
/note="assembly_fragment"
85763..97087
/note="assembly_fragment"
97188..113019
/note="assembly_fragment"
113120..130955
/note="assembly_fragment"
131056..148692
/note="assembly_fragment"
148793..165067
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
165168..185335
/note="assembly_fragment"
185436..202437
/note="assembly_fragment"
  
```

ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 202437;
 Best Local Similarity 94.7%; Pred. No. 1.1e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19

Db 171824 CGAGTGCCAGCCCCACA 171806

RESULT 27

RME603643 RME603643 260050 bp DNA linear BCT 05-JUL-2002
 LOCUS Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymb;
 DEFINITION segment 2/6.

ACCESSION AL603643 AL591985

VERSION AL603643.1 GI:15140164

KEYWORDS

SOURCE Sinorhizobium meliloti (Rhizobium meliloti)

ORGANISM Sinorhizobium meliloti
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

REFERENCE

AUTHORS

Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P.,
 Vorholter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J.,
 Golding, B. and Puhler, A.
 The complete sequence of the 1,683-kb pSymb megaplasmid from the
 N2-fixing endosymbiont Sinorhizobium meliloti

Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)

21396508

11481431

epub ahead of print

2 (bases 1 to 260050)

Weidner, S.

AUTHORS

TITLE

JOURNAL

MEDLINE

REMARK

Submitted (07-JUN-2001)

Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
 Germany

| | | | |
|----------|---|--|--|
| COMMENT | Submitted on behalf of Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1 Canada mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE PEXO, pSymB. | | |
| FEATURES | Location/Qualifiers 1..260050 /organism="Ginorhizobium meliloti" /mol_type="genomic DNA" /strain="1021" /db_xref="taxon:382" /plasmid="pSymB" complement(120..281) /gene="SMB20302" complement(120..281) /genes="SMB20302" /function="MISCELLANEOUS; Hypothetical/Global homology" /note="Product confidence : putative" Gene name confidence : hypothetical /codon_start=1 /evidence=not_experimental /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAC48692.1" /db_xref="GI:15140165" /db_xref="GOA:Q92WN7" /translation="MSVLISILITFLWLVLVLYLVNRLPLDGRTRKQIVQVIVIIIGIL SLRLYLAVP" 478..705 /gene="SMB20303" 478..705 /genes="SMB20303" /function="MISCELLANEOUS; Unknown" /note="Product confidence : hypothetical" Gene name confidence : hypothetical /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOTHETICAL PROTEIN" /protein_id="CAC48693.1" /db_xref="GI:15140166" /db_xref="GOA:Q92WN6" /db_xref="SPTREMBL:Q92WN6" /translation="MSVRFSDLAIDDDITLMLRRVHVYACSSRELEPSEHGAEIGKM LFLHYNQVRSRLALMQLTAGAAPPAEAG" complement(746..1804) /note="this element seems to be partial or inactive" /insertion_seq="ISRm2011-2/ISRm11 OR SMB21616" complement(762..1367) /genes="TRm2011-2C OR SMB20304" complement(762..1367) /genes="TRm2011-2C OR SMB20304" /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions" /note="Product confidence : probable" Gene name confidence : putative /codon_start=1 /evidence=not_experimental /transl_table=11 /product="probable ISRm2011-2 transposase protein, C-terminal portion" /protein_id="CAC48694.1" /db_xref="GI:15140167" /db_xref="GOA:Q92527" /db_xref="SPTREMBL:Q92527" /translation="KTVLSRERPDVARHARWLKHCPIIDPARLVFIDETWTKTN MAPLRWARGRSELVGYAPFGWNTMTFVAALRADRVSAPIIDGPIINGERFRLVQQ VLVPEUKAGDI VILNDLIGSHKGEINAAIRKAGARLFFDKYSPDLNFIKLFPAKIKH WLREQAQRSDAIHDLRHILQAVTFQECAYFKRAGYERA" 1002..1373 /genes="TRm2011-2 OR SMB20306" /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions" /note="Product confidence : putative" Gene name confidence : putative /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative TRm2011-2 protein" /protein_id="CAC48695.1" /db_xref="GI:15140168" /db_xref="GOA:Q52888" /db_xref="SPTREMBL:Q52888" /translation="MGAEIVQDDVDVAGFFRQNLIDIAEAFADRAIEDKGSADAV GPECGDGHGVPMAGVAGVAFQFAARRPAPQRHVRRLPGLIDENAGGINSAGVLQP SCFVAGDVGLPPAGDQRLFL" complement(1301..1708) /genes="TRm2011-2N OR SMB20305" complement(1301..1708) /genes="TRm2011-2N OR SMB20305" /function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related functions" /note="Product confidence : probable" Gene name confidence : putative /codon_start=1 /evidence=not_experimental /transl_table=11 /product="probable ISRm2011-2 transposase protein, N-terminal portion" /protein_id="CAC48696.1" /db_xref="GI:15140169" /db_xref="GOA:Q926A8" /db_xref="SPTREMBL:Q926A8" /translation="MARPFSNDLRERVDVAVTGESLSCRAAAKRGIGISTALDWVR FRETSAAPQMGHGRPKRLSGPHRAWLLCRERDFTLHGLVAELSERGLKVDYNAV WTFVEEGLSYKKRWSPPANGSGFTSPATGHDG" complement(1801..2448) /genes="SMB20307" complement(1801..2448) /genes="SMB20307" /function="MISCELLANEOUS; Hypothetical/Partial homology" /note="Product confidence : putative" Gene name confidence : hypothetical /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative dihydroxyacetone kinase protein" /protein_id="CAC48697.1" /db_xref="GI:15140170" /db_xref="GOA:Q92WN5" /db_xref="SPTREMBL:Q92WN5" /translation="MMSESAVRRLIEVFSAISAIEARDHLSLSDGAIGDADHGITM AIGFSAVTGLARTDAGSLGVTAAASAFNLNAVAGSTGPLYATAFRAAQALEGR DTLIDACSLIQAASGIGERGKGORGDKMLDWLPAADAAQAATAAKSTAAFWA DVTAAENGASATRSVATKGRARVGRSLGHMDPGAASAVLIIRAAQTLLREG" complement(2527..4614) /genes="SMB20312" complement(2527..4614) /genes="SMB20312" /function="MISCELLANEOUS; Not classified regulator" /note="Product confidence : putative" Gene name confidence : hypothetical /codon_start=1 /evidence=not_experimental /transl_table=11 /product="putative transcriptional regulator protein" /protein_id="CAC48698.1" /db_xref="GI:15140171" /db_xref="GOA:Q92WN4" /db_xref="SPTREMBL:Q92WN4" /translation="MTIKTSARKGAANSRGTIDASDAIPLYRGDDPFYVWACHLYYED GMTQGEIAEANGVSRATVNSYLADAREKGINVISIEPARLASUTIIAQELRRHFLTDC | | |

```

LVPSDDNARPLIDRLGAGAGQALPKLKSGDGLAVANGRTVLVSGEHAGASLDQMT
VVOAGGTTASPAYTEPCASARAISARCVNITAPAVKSPVLLHMLLEPVOEQ
FATLARNVFGISLRFNSITHTSGFFSVSLQDYLAAGAVGVAGRFDERGRPV
AGPLDRTVGISDLRLKIGTRIAVAGGDKVPALLAALRGYVNVLTDAATGHGIL
RADGVTALDKLWHRKLPVTPSSVYRTHVKFPLNPNVWEEMLDGVAKAHEIYLQPI
NGSHRALVARNRPRPKVGLVIGGCTGHEPCFLGVGVGKLADAVAGNI FSSPPDPPI
VQCAASGSGEVLVYGNVAGDVMPKMAELAEKIGIKVKTVLTTDDVAVSVEER
BGRGVAGNFFYFKIAGACDRGLSEACEAVIRKANURKTYTVGVALEPGSPQTRRH
NFEIGDDMEVGHGIFGVTRRIRSADEITDSIMDRI FKEMKTAPEGVAVLVNS
FGATLMELYLILFRVQORLAADKILIEANWIGHYCTSLDMAGASITILHLDQELSDL
LHHPCTAPLKNV"
complement(4728, .5369)
/gene="Smb20313"
/gene="Smb20313"
complement(4728, .5369)
/gene="Smb20313"
/function="Small Molecule Metabolism; Central intermediary
metabolism; misc. glycerol metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative dihydroxyacetone kinase protein"
/protein_id="CAC48699.1"
/db_xref="GI:15140172"
/db_xref="GOA:O92WN3"
/db_xref="SPTREMBL:O92WN3"
/translation="MQTFENARAGDIVTLADRIENRAYLSEIDGKIGDGHGVNWA

Query Match      87.0%; Score 17.4; DB 1; Length 260050;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CGAGGCGCCGACCCCA 19
      |||||
Db      1207 CGAGGCGCCGACCCCG 1225

RESULT 28
AE006467
LOCUS      Homo sapiens 16p13.3 sequence section 6 of 8.
DEFINITION
ACCESSION      AE006467
VERSION      AE006467.1 GI:14336745
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 279011)
    Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,
    Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and
    Higgs,D.R.
    Sequence, structure and pathology of the fully annotated terminal 2
    Mb of the short arm of human chromosome 16
    Hum. Mol. Genet. 10 (4), 339-352 (2001)
  2 (bases 1 to 279011)
    Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,
    Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and
    Higgs,D.R.
    Direct Submission
    Submitted (07-DEC-2000) MRC Molecular Haematology Unit, Weatherall
    Institute of Molecular Medicine, John Radcliffe Hospital, Oxford,
    Oxon OX3 9DS, UK
    Location/Qualifiers
      1..279011
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="16"
        /map="16p13.3"
FEATURES
      source
        1..14033
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 358b7"
        1833..46356
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 316g12"
        39209..78044
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 312e8"
        95097..126950
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 390e8"
        142264..185653
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 305c8"
        184733..220367
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 385e7"
        187817..235498
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 380f5"
        232370..276380
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 313f4"
        248926..279011
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 425c2"
        275910..279011
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="cosmid 315g5"
        <1..1932
          /note="HS358B7"
          /complement(1294, .1450)
          /note="MIR; RepeatMasker predicted
            23/10/2000"
          /rpt_family="SINE/MIR"
          1833..39308
            /note="HS316G12"
            2453..2501
              /note="MER5A; RepeatMasker predicted
                23/10/2000"
              /rpt_family="DNA/MER1_type"
              2517..2817
```

us-10-624-714-11.rge

Mon Jun 21 09:02:04 2004

```

/note="AluJo; RepeatMasker Predicted
23/10/2000"
/rpt_family="SINE/Alu"
/note="MER5A; RepeatMasker Predicted
23/10/2000"
/rpt_family="DNA/MER1_type"
2839_4637
/note="CpG island; BAIAP3"
/evidence=not_experimental
3444_3502
/note="GC rich; RepeatMasker Predicted
23/10/2000"
/rpt_family="Low_complexity"
3559_19342
/genes="BAIAP3"
/note="synonym: BAP3"
4319_4339
/note="GC rich; RepeatMasker Predicted
23/10/2000"
/rpt_family="Low_complexity"
join(4565_4659,8443_8583,8805_8892,9067_9147,
9399_9506,10825_10884,11032_11098,11197_11392,
11725_11769,11889_11966,12084_12186,12430_12537,
12622_12717,12816_12919,13296_13416,13925_14049,
14130_14220,14372_14428,14504_14601,14701_14752,
14872_15002,15152_15275,15650_15793,15889_15981,
16073_16264,16345_16413,16508_16587,16750_16858,
17201_17272,17448_17528,17604_17724,17797_17920,
17999_18204,18282_18385)
/genes="BAIAP3"
/note="Similar to Unigene cluster Hs.101516"
join(4565_4659,8443_8583,8805_8892,9067_9147,
9399_9506,10825_10884,11032_11098,11197_11392,
11725_11769,11889_11966,12084_12186,12430_12537,
12622_12717,12816_12919,13296_13416,13925_14049,
14130_14220,14372_14428,14504_14601,14701_14752,
14872_15002,15152_15275,15650_15793,15889_15981,
16073_16264,16345_16413,16508_16587,16750_16858,
17201_17272,17448_17528,17604_17724,17797_17920,
17999_18204,18282_18385)
/genes="BAIAP3"
/codon_start=1
/product="BAI associated protein 3"
/protein_id="AAK61275.1"
/db_xref="GI:14336746"
/db_xref="MIM:604009"
/db_xref="SPTREMBL:O94812"
/translacion="MPEKGAFAAGPPDLHGTALGPAGIWRSRSPAMSTLLDIKS
SVLRQVQCFRRFEQPGSASADPOEPATGAMKPGDGEFFAHMLMLKKGSGRO
GLPCLEVRSGPAPPEPDSLSGLRALAPPEVMEALYTVLYAGTNGPQVD
DEALLSYQQVFGSLPEEETAEIRKAPATKALKVSWRANLAKDPNGSDP
YCMGLIPASDATREAPKQRFGRKSGPLPAKCIQIVTKVSTLNPMWKEH
FLPEIEDVSDQLHLDIWDHDDVSLVEAKRNLVIGLKGMYFKQIVKSARANGT
AGPTEDDFGLCLMLPVREVPVAGVDFWFKLEPRSSASRVQGHCHLVLLITQFD
TANSORGRSGFTSHLILLSHLLRLEHSAEPNSNWRGELSTPAATILCHGAQNSLS
PQLQVLHVQVSRRHQCTLDLYSYLLGLLEDQMAHWEAPSLPQOESELSAF
SEFGLQLRLQDRFPATNSAVHRLLEKLKQLQLPQPSFEFCPESELSNMDIAA
ALKRKNREYRLNDKSPREPQGPORPLGLVLADAVDDQFCISVIALFSLIN
VDVFTTLFQLERLVAEEAWLTELSPKTVLEASGLFELYTLADQSRWDSIPGR
DSLSLALAGHAFPLAVKLWFQVLRQAKRLOQAVMDTLEPVDASRHSSTAATA
GLCLSHQELVWFLAPDAQGLGTQGVQCVETATLYTELLKRVDTOPGAAGEA
VSPALCVNLNVLKPAQAKGLAWPEGATGEGVLPRLPSTCTOALDLOREA
HTYHALTSKWGCDIKYVQHLSLSPDSIONDEAVAPLMKYLDELKALINASLGKLN
SVRYELWELLQALQALCANRVDSDYFSPHFLEALVFFAEAGGGLPLESLRD
GSYKRLKBEELRLQCTRECIEQFYLDKLRQKRLQNRGLSVSCHEAEARQIAVE
VLAHADLLPLDANGLSDPFIYVELGPPHFLPVRQRTQVKTTRTLHPVYDELFYFVP
AEACRRRAACVLTVMHDWLTNDPFAAGALGLGVTGVARPQVGGGARQGPVTLH
LCRPRAQVARSARLRLSGTSKEAQEFVKLEKLEKMEADP"
5744_5876
/note="MIR; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/MIR"
7123_7134
/note="AluSc; RepeatMasker Predicted
23/10/2000"
/rpt_family="SINE/Alu"
7135_7385
/note="AluSq; RepeatMasker Predicted
23/10/2000"
/rpt_family="SINE/Alu"
7386_7478
/note="AluSc; RepeatMasker Predicted
23/10/2000"
/rpt_family="SINE/Alu"
7572_7592
/note="(CTG)n; RepeatMasker Predicted
23/10/2000"
/rpt_family="Simple_repeat"
9646_9787
/note="(CA)n; RepeatMasker Predicted
23/10/2000"
/rpt_family="Simple_repeat"

Query Match 87.0%; Score 17.4; DB 9; Length 279011;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGAGCCCA 19
|||||
DB 8592 CGAGGCGCCCGAGCCCA 8610

RESULT 29
SNES91785 286550 bp DNA linear BCT 05-JUL-2002
LOCUS SNES91785
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 4/12.
ACCESSION AL591785 AL591688
VERSION AL591785.1 GI:15073719
KEYWORDS
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 286550)
Capela,D., Barloy-Hubier,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kias,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ransperger,U.,
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
21396507
MEDLINE 11481430
PUBMED 2 (bases 1 to 286550)
Gouzy,J.
Direct Submission
Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
MELILO EU Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR8061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@oulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
Location/Qualifiers
1. 286550
source

```

```

/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
70..813
/gene="gst11 OR SMC00916"
70..813
/gene="gst11 OR SMC00916"
/EC_number="2.5.1.18"
/function="miscellaneous"
/notes="Product confidence : putative
Gene name confidence : putative
Predicted by Codon usage
Predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/db_xref="GOA:Q92RR5"
/db_xref="SPTREMBL:Q92RR5"
/translation="MADLSRPITARWPAKNPDIILQYSLPTNGVKVSIAEELGLA
YEAHRISEPTNEQKSPFELSINPGRIPALIDPNPDKPIGLFSGAILLYLAETKG
KLUPDAARRYETLCWVFMFGMGQPGHFFKFAADVANNPYVQRYDEAKR
LUGVLEAREGKWLWGGNTPTSPTRIPGKRGSSWVAPRCSTIRASRSWTGW
AAWHLRPRARVRSRARSRLSPAGD"
588..779
/gene="SMC00915"
588..779
/gene="SMC00915"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/db_xref="SPTREMBL:Q92KL5"
/translation="WADGGEYTIADIATYPWEGARFYGGAELVDYKSPFNWDVW
DRGLRPAAPAGQMEIPRKE"
867..2084
/gene="SMC00914"
867..2084
/gene="SMC00914"
/function="small molecule metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC45363.1"
/db_xref="GI:15073722"
/db_xref="GOA:Q92RR4"
/db_xref="SPTREMBL:Q92RR4"
/translation="MSGLVVGQGAAPALVAKLRKDKRPIITVVAASLPYQRP
PLSKYLREMLDRLRLTPRPAWYAEHIDIRLSTTVTRDLRAQVALSDGSMLTTE
TLAATGTPRLPAVGGDLGVVVRDFDADRLAEQPGSRVIVVGGYIGLEA
AAVATSGLEVTVEMADRIQVRASAATSARVIRHSRQVDIRERTGLRHIGDGA
RVTAELSDGSPVFDIVIGVIGVAAANDALAEAGIETANGIVVDVSHRSTDPITVAM
GDCAVLPMWGMRIKRLSVONADQAEAAVAGGTDPDPKPMFWSQDQVVKLIQIAG
FGLGHDETFLVQGORQGSVSVYFRQKLIIVDAINDAKAYVTGCKLLEAGATPDRLL
LADQVDLKAIIY"
complement(2114..2572)
/notes="Sm-2 OR SMC04876
REPEAT SM-2
Predicted by Homology"
/evidence=not experimental
complement(2585..4222)
/gene="groEL1 OR SMC00913"
complement(2585..4222)
/gene="groEL1 OR SMC00913"
/function="cell processes; chaperoning"
/notes="Predicted by Codon usage
Predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="60 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45364.1"
/db_xref="GI:15073723"
/db_xref="GOA:P35469"
/db_xref="SWISS-PROT:P35469"
/translation="MAAKEVYKFSAREKMLRGVDILADAVKVTLPKGRNVVIDKSP
GAPRITKDGVSVAKEIELEDKFENMGQWVREVASKNTNDIAGDGTITATVLAQIVRE
GAKAVAGMFMOLKRGIDJLAVAEVVDLLAKKAKKNTSDEVAQVGTISANGKQIGL
DIABAMQKNGNEGVIIVTEAKTAELEVEVQGFDRGYSLPYVTNPERVADLEDA
FIILHEKLSNLOAMLPVLEAVVQTKPLIIAEDVEGEALTLVWNLKGGIKIAAV
KAPGFGRRKAMLEDAIITGGTVIISDGLIILESVTDLMLGRKAVIKENTTIVD
GAGQSDIEGRVAQIAQIETTSYDREKLOERLAKGAVIRVGGATEVEVKEK
KDRIDDAIATRAVQEGIVPGGGVALLSSVKITVKGENDDQDAGVIVRRALQSPA
RQIVENAGDEASIVVGKILSKNTDDFGYNAQTGEYGDMTIANGIIDPVKVVRTALQDAA
SVASLLITTEAMIAELPKDAPAMPGGMGGMGMDMM"
complement(4298..4594)
/gene="groES1 OR SMC00912"
complement(4298..4594)
/gene="groES1 OR SMC00912"
/function="cell processes; chaperoning"
/notes="Predicted by Codon usage
Predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="10 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45365.1"
/db_xref="GI:15073724"
/db_xref="GOA:P35473"
/db_xref="SWISS-PROT:P35473"
/translation="MASTNFRPLHDRVVRVVESEKTKGIIIPDTAKEKPOEGEIV
AVGSGARDSGKVVPDLVDKAGDRILFGKSGTEVKINGEDLLIMKEADIMGVIG"
4863..5690
/gene="SMC00911"
4863..5690
/gene="SMC00911"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CONSERVED HYPOTHEITICAL PROTEIN"
/protein_id="CAC45366.1"
/db_xref="GI:15073725"
/db_xref="SPTREMBL:Q92RR3"
/translation="VASFNSVISEKLARLVGTFKAPVIVDVRDDDDFAADHHLVPGSF
HKGHVHPQWAEYRGWVYVLCIKHGKLGSEGAALWLRHAGAAAILDGGIEMWRAAG
LPIVPAFNPSPDPAGSVVWVRSRPFKIDRIACPWLIRFVDPFRARFLVTPAEVEAV
GERFGATPDIDEGVFWSHRGECTFTMTYKESGLSPALEHVARIVRGADTDRRELEP
EAAGLLVASLGLSRMYADDLLEAGMTIYDALYWARDATQEKHDVWSHGKGRKP"
6004..6852
/gene="SMC00910"
6004..6852
```

Mon Jun 21 09:02:04 2004

```

/gene="SMC00910"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC45367.1"
/db_xref="GI:15073726"
/db_xref="GOA:Q92RR2"

Query Match      87.0%   Score 17.4; DB 1; Length 286550;
Best Local Similarity 94.7%; Pred. No. 1.le+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCGCCACA 19
Db 25466 CGAGGCGCCAGCCGCCCA 25484

RESULT 30
SME591785/c
LOCUS SME591785 286550 bp DNA linear BCT 05-JUL-2002
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 4/12.
ACCESSION AL591785 AL591688
VERSION AL591785.1 GI:15073719
KEYWORDS
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 286550)
Boisard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,J.,
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
2 (bases 1 to 286550)
Gouzy,J
Direct Submission
Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
MELILO EU Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Génétique et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Genetique Physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
Location/Qualifiers
1. 286550
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
70. .813
/gene="gst11 OR SMC00916"
70. .813

FEATURES
source
1. 286550
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
70. .813
/gene="gst11 OR SMC00916"
70. .813

/gene="SMC00916"
/EC number="2.5.1.18"
/function="miscellaneous"
/notes="Product confidence : putative
Gene name confidence : putative
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/db_xref="GOA:Q92RR5"
/db_xref="SPTREMBL:Q92RR5"
/transl_table="MADLASPITATAPAKNPDIOLYSLPTPGVGVKVSIALBELGLA
YEHRISGPDNTEQKSPFSLNPGRIPIALIDPNGDGPGLIGLFGSGATLIYLAEXTG
KLIPADARRYETLCWFMFGMIGFNGFGHFFKFAADKXANNPPYQVRYDEAKR
LLGVLEARLEGQWLNGGRNTPSPTSRPIPIPMKRGSSWVAFRCSTIRASRWSTGWI
AAWHALRRPRAWRSRARSRSRSLPAD"
588. .779
/gene="SMC00915"
588. .779
/gene="SMC00915"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/db_xref="SPTREMBL:Q92KL5"
/transl_table="MADGGEYTIADITYPWEGARKFYGGAEVLDYKSFNPMVMDV
DRGLARPAQKQMEIPKKE"
867. .2084
/gene="SMC00914"
867. .2084
/gene="SMC00914"
/function="small molecule metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC45363.1"
/db_xref="GI:15073722"
/db_xref="GOA:Q92RR4"
/db_xref="SPTREMBL:Q92RR4"
/transl_table="MSCRIYVGGQAAFAVAKLRALKDMRPITVVAEASLPYQRP
PLSKYLIREMTLDLLRYEAKVAEHEIDRLSTVTRVDRLAKOVALSDGSMLYTVE
TLAPATGATPRLPAVGGDLGAVGVVVRDRDRLAEEMQGRVILVGGYIGLEA
AAVATSGLEVTVIEMADRIQORVASAATSAIYREIHRSGVDIRERTGHLRIQDNG
RVTAELSDGSVIPDIVIGVANDALAHAGIETANGIYVDSHGRTSDFTIVAM
GDCVLPWDGMRIRLESQVNAQAEAAVLAGTDPDPKPFWSDDQYDKVLQIAG
FGLGDETLVROGQGSVSVVYPRQGLIADVAINDAKAYVTGKKLLEAGATPDRLL
LADQVLDKALLV"
complement(2114. .2572)
/notes="Sm-2 OR SMC04876
REPEAT SM-2
Predicted by Homology"
/evidence=not_experimental
complement(2585. .4222)
/gene="groEL1 OR SMC00913"

```

CDS

complement(2585..4222)
/gene="groEL1 OR SMC00913"
/function="cell processes; chaperoning"
/note="predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental

/transl_table=11
/product="60 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45364.1"
/db_xref="GI:15073723"
/db_xref="GOA:P35469"
/db_xref="SWISS-PROT:P35469"

/translation="MAAEVVFGRSAREKMLRGVLDLADAVKVTILGPKRNVWIDKSP
GAPRITKQCVSAKEIELEDKFNMGAGQWREVASKTNDIAGDQTTTATVLAQAIURE
GAKAVAGMNPMDLARGIDILVAEVVDKLLAKAKINTSDEVAOVGTISANGKQIGL
DIAEAMQKVGNEGVITVEAKTAETELVWGMQDRGILSPYFVTFNPEKMWALIEDA
FILLHEKLSNLQMLPVLVAEVVQIGKELLIAEDVEGALATLVNKLGLKIAAV
KAPGFGRRKAVLEDAIILTGTVISEDLGKLESVTILDMGLGRKKVSIITENTIVD
KAGQSDTEGRVAQIAQTBEFTSDYDREKLQERLAKLAGGAVIRVGGATEVEVKEK
KDRIDNALNATRAAQEGTIPGGVALLRSVKITVKGENDQDQAGVNVIRALQSPA
ROI VENVAGDEASIVVGGKILEKNTDGFVNAOTGEYGDMIANGIIDPVKVVRTALQDAA
SVASLLITTEAMIAELPKKAPAMPGGMGGMDM"
complement(4298..4594)

gene

/gene="groES1 OR SMC00912"
complement(4298..4594)

CDS

/gene="groES1 OR SMC00912"
/function="cell processes; chaperoning"
/note="predicted by Codon_usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not_experimental

/transl_table=11
/product="10 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45365.1"
/db_xref="GI:15073724"
/db_xref="GOA:P35473"
/db_xref="SWISS-PROT:P35473"

/translation="WASNFRPLHDRVVRVVESEKTKGGIIPDTAKEKPOGEIV
AVGSGARDESGKVPFLVDYKAGDRILFGKMSTEVKINGEDLLIMKEADIMGVIG"
4863..5690

gene

/gene="SMC00911"
4863..5690

CDS

/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD"
/codon_start=1
/evidence=not_experimental

/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC45366.1"
/db_xref="GI:15073725"
/db_xref="SPTREMBL:Q92RR3"

/translation="WASFNSISSEKLARLVCTPKAPVIVDVRDDDDFAADHHLVPGSF
IRGHVHVQWAEETRGVWVVICVHGKLGSEGAANLHAGAAEILDGGLEAWRAAG
LPLVVASLNSDPAGSGVWVVRSPKIDRIACPLWIRFVDPVDFRFLYVTPAEVAV
GERGATFDIEGVFWSHRGECTFDTNWKEFGLSFPALHEVARIVRGADTDRELEP
EAGGLAVSLGLSRMYADDLQLEAGMTIYDALYRWARDAQEKXHDWVSRGKGKRP"
6004..6852

gene

/gene="SMC00910"
6004..6852

CDS

/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"

/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC45367.1"
/db_xref="GI:15073726"
/db_xref="GOA:Q92RR2"

Query Match 87.0%; Score 17.4; DB 1; Length 286550;
Best Local Similarity 94.7%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGGCCCCAGCCCCACA 19
Db 284901 CGAGGGCCCCAGCCCCGCA 284883

Search completed: June 20, 2004, 11:42:05
Job time : 569.069 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 114.921 seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-11
Perfect score: 20
Sequence: 1 cgaggccgcccagcccccacaa 20

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : N Geneseq 29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 19 | 95.0 | 110000 | 4 | AAI99682_25 |
| 2 | 19 | 95.0 | 110000 | 4 | AAI99682_26 |
| 3 | 19 | 95.0 | 110000 | 4 | AAI99683_25 |
| 4 | 16.8 | 84.0 | 410 | 2 | AXX20696 |
| 5 | 16.8 | 84.0 | 993 | 6 | ABQ75352 |
| 6 | 16.8 | 84.0 | 1002 | 6 | AAQ73099 |
| 7 | 16.8 | 84.0 | 1830 | 7 | ACA52540 |
| 8 | 16.8 | 84.0 | 9851 | 9 | ADC86884 |
| 9 | 16.8 | 84.0 | 19186 | 2 | AXX20620 |
| 10 | 16.8 | 84.0 | 21501 | 9 | ADC35539 |
| 11 | 16.8 | 84.0 | 28000 | 9 | ADD00956 |
| 12 | 16.8 | 84.0 | 37113 | 6 | AAQ75121 |
| 13 | 16.4 | 82.0 | 354 | 3 | AAQ10189 |
| 14 | 16.4 | 82.0 | 3145 | 3 | AAQ59241 |
| 15 | 16.4 | 82.0 | 3287 | 6 | AAQ59902 |
| 16 | 16.4 | 82.0 | 3345 | 6 | ABN59766 |
| 17 | 16.4 | 82.0 | 19951 | 7 | ABZ24596 |
| 18 | 16 | 80.0 | 3417 | 6 | ABV78124 |
| 19 | 16 | 80.0 | 3417 | 6 | ABZ35700 |
| 20 | 16 | 80.0 | 3417 | 6 | ABX03943 |
| 21 | 16 | 80.0 | 3417 | 6 | ABL91665 |
| 22 | 16 | 80.0 | 3845 | 2 | AAQ47064 |
| 23 | 16 | 80.0 | 3845 | 6 | ABZ34892 |

| | | | | | | | | | | | | | |
|-----|------|------|------|---|----------|--------------------|-----|------|------|------|---|----------|--------------------|
| 97 | 15.4 | 77.0 | 858 | 6 | ABK99585 | Abk99585 Rat proba | 170 | 15.2 | 76.0 | 448 | 2 | AAQ59153 | AaQ59153 Human bra |
| 98 | 15.4 | 77.0 | 858 | 7 | ACD07310 | AcD07310 Human muc | 171 | 15.2 | 76.0 | 475 | 7 | ACC44118 | Acc44118 Herpes si |
| 99 | 15.4 | 77.0 | 882 | 2 | AX57407 | Ax57407 Rat U3 ge | 172 | 15.2 | 76.0 | 523 | 4 | AAH21061 | AaH21061 Parallel |
| 100 | 15.4 | 77.0 | 984 | 6 | AB27499 | Ab27499 Human G-p | 173 | 15.2 | 76.0 | 665 | 7 | ABT40532 | AbT40532 Toxicity |
| 101 | 15.4 | 77.0 | 984 | 6 | AB242916 | Ab242916 Human GFC | 174 | 15.2 | 76.0 | 903 | 7 | ACF39423 | AcF39423 Mycobacte |
| 102 | 15.4 | 77.0 | 984 | 6 | AD61656 | Ad61656 Human GFC | 175 | 15.2 | 76.0 | 906 | 7 | ACA40706 | AcA40706 Prokaryot |
| 103 | 15.4 | 77.0 | 990 | 6 | AAI99826 | AaI99826 Human G p | 176 | 15.2 | 76.0 | 963 | 7 | ACA39806 | AcA39806 Prokaryot |
| 104 | 15.4 | 77.0 | 990 | 6 | ABV72908 | AbV72908 Human G p | 177 | 15.2 | 76.0 | 1126 | 6 | ABS64603 | AbS64603 Human cDN |
| 105 | 15.4 | 77.0 | 990 | 6 | AL54452 | Al54452 Human G p | 178 | 15.2 | 76.0 | 1146 | 6 | ACC78418 | Acc78418 E. desert |
| 106 | 15.4 | 77.0 | 993 | 6 | AB154463 | Ab154463 Human GFC | 179 | 15.2 | 76.0 | 1449 | 9 | ACC78413 | Acc78413 E. desert |
| 107 | 15.4 | 77.0 | 993 | 6 | ABK91022 | AbK91022 Human G-p | 180 | 15.2 | 76.0 | 1796 | 9 | ADC64479 | AdC64479 Human inh |
| 108 | 15.4 | 77.0 | 993 | 6 | ABK91022 | AbK91022 Human GFC | 181 | 15.2 | 76.0 | 1806 | 6 | AAH77885 | AaH77885 Nucleotid |
| 109 | 15.4 | 77.0 | 993 | 6 | AAI51747 | AaI51747 Human G p | 182 | 15.2 | 76.0 | 1806 | 6 | ABK78428 | AbK78428 Human cDN |
| 110 | 15.4 | 77.0 | 993 | 8 | ACC85034 | Acc85034 Human GFC | 183 | 15.2 | 76.0 | 1806 | 6 | AAH77885 | AaH77885 Nucleotid |
| 111 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human G p | 184 | 15.2 | 76.0 | 2426 | 1 | AAH77885 | AaH77885 Nucleotid |
| 112 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 185 | 15.2 | 76.0 | 2426 | 1 | AAH77885 | AaH77885 Nucleotid |
| 113 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 186 | 15.2 | 76.0 | 2784 | 7 | ACA37694 | AcA37694 |
| 114 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 187 | 15.2 | 76.0 | 2839 | 9 | ADC30721 | AdC30721 Human nov |
| 115 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 188 | 15.2 | 76.0 | 2839 | 9 | ADC30721 | AdC30721 Human nov |
| 116 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 189 | 15.2 | 76.0 | 2839 | 9 | ADC30721 | AdC30721 Human nov |
| 117 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 190 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 118 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 191 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 119 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 192 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 120 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 193 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 121 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 194 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 122 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 195 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 123 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 196 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 124 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 197 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 125 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 198 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 126 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 199 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 127 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 200 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 128 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 201 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 129 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 202 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 130 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 203 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 131 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 204 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 132 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 205 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 133 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 206 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 134 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 207 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 135 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 208 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 136 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 209 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 137 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 210 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 138 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 211 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 139 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 212 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 140 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 213 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 141 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 214 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 142 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 215 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 143 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 216 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 144 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 217 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 145 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 218 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 146 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 219 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 147 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 220 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 148 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 221 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 149 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 222 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 150 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 223 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 151 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 224 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 152 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 225 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 153 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 226 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 154 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 227 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 155 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 228 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 156 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 229 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 157 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 230 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 158 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 231 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 159 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 232 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 160 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 233 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 161 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 234 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 162 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 235 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 163 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 236 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 164 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 237 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 165 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 238 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 166 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 239 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 167 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 240 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 168 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 241 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |
| 169 | 15.4 | 77.0 | 1038 | 7 | AB242594 | Ab242594 Human REM | 242 | 15.2 | 76.0 | 3016 | 4 | AAH16088 | AaH16088 Human cDN |

| | | | | | | |
|-------|------|------|-----|---|-----------|-----------|
| 243 | 14.8 | 74.0 | 238 | 3 | AAC28264 | Human sec |
| C 244 | 14.8 | 74.0 | 240 | 4 | AS36090 | Human car |
| C 245 | 14.8 | 74.0 | 240 | 9 | AD46784 | Human car |
| C 246 | 14.8 | 74.0 | 247 | 7 | AD48755 | Human acy |
| C 247 | 14.8 | 74.0 | 253 | 4 | AS36088 | Human car |
| C 248 | 14.8 | 74.0 | 253 | 4 | AS36087 | Human car |
| C 249 | 14.8 | 74.0 | 253 | 9 | AD46782 | Human car |
| C 250 | 14.8 | 74.0 | 253 | 9 | AD46781 | Human car |
| C 251 | 14.8 | 74.0 | 316 | 2 | AT22911 | Human gen |
| C 252 | 14.8 | 74.0 | 332 | 2 | AC60504 | Human bra |
| C 253 | 14.8 | 74.0 | 332 | 6 | ABL85098 | Human ova |
| C 254 | 14.8 | 74.0 | 353 | 6 | ABL84737 | Human ova |
| C 255 | 14.8 | 74.0 | 389 | 4 | AA123824 | Human bre |
| C 256 | 14.8 | 74.0 | 391 | 6 | ABL87076 | Human ova |
| C 257 | 14.8 | 74.0 | 422 | 4 | AI92169 | Human pol |
| C 258 | 14.8 | 74.0 | 426 | 4 | AA129034 | Colon tum |
| C 259 | 14.8 | 74.0 | 426 | 7 | AB233220 | Human col |
| C 260 | 14.8 | 74.0 | 438 | 4 | AA114987 | Probe #49 |
| C 261 | 14.8 | 74.0 | 438 | 4 | ABA56720 | Human foe |
| C 262 | 14.8 | 74.0 | 438 | 4 | AI36333 | Probe #50 |
| C 263 | 14.8 | 74.0 | 438 | 4 | ABA46177 | Human bre |
| C 264 | 14.8 | 74.0 | 438 | 4 | ABA26339 | Probe #48 |
| C 265 | 14.8 | 74.0 | 438 | 4 | AK30378 | Human bon |
| C 266 | 14.8 | 74.0 | 438 | 4 | AK04850 | Human bra |
| C 267 | 14.8 | 74.0 | 438 | 4 | AB330017 | Human liv |
| C 268 | 14.8 | 74.0 | 438 | 5 | AA104752 | Probe #47 |
| C 269 | 14.8 | 74.0 | 438 | 6 | AB304976 | Human gen |
| C 270 | 14.8 | 74.0 | 474 | 2 | AAV86659 | EST clone |
| C 271 | 14.8 | 74.0 | 500 | 4 | AA114971 | Human bre |
| C 272 | 14.8 | 74.0 | 501 | 8 | ACH31903 | Human end |
| C 273 | 14.8 | 74.0 | 502 | 4 | AA198936 | Human exc |
| C 274 | 14.8 | 74.0 | 502 | 5 | AA163286 | Human kid |
| C 275 | 14.8 | 74.0 | 519 | 4 | AAH13057 | Human cdn |
| C 276 | 14.8 | 74.0 | 520 | 6 | ABN84933 | Human glu |
| C 277 | 14.8 | 74.0 | 532 | 7 | AB236649 | Human gen |
| C 278 | 14.8 | 74.0 | 532 | 3 | AB250398 | Human end |
| C 279 | 14.8 | 74.0 | 565 | 4 | AA117173 | Probe #71 |
| C 280 | 14.8 | 74.0 | 565 | 4 | ABA61700 | Human foe |
| C 281 | 14.8 | 74.0 | 565 | 4 | AA141612 | Probe #10 |
| C 282 | 14.8 | 74.0 | 565 | 4 | ABA29343 | Probe #78 |
| C 283 | 14.8 | 74.0 | 565 | 4 | AK35896 | Human bon |
| C 284 | 14.8 | 74.0 | 565 | 4 | AK10002 | Human bra |
| C 285 | 14.8 | 74.0 | 565 | 4 | AB35605 | Human liv |
| C 286 | 14.8 | 74.0 | 565 | 6 | AB310119 | Human gen |
| C 287 | 14.8 | 74.0 | 591 | 6 | AAH10551 | Human cdn |
| C 288 | 14.8 | 74.0 | 591 | 6 | ABN84931 | Human glu |
| C 289 | 14.8 | 74.0 | 596 | 4 | AAH69170 | Human cer |
| C 290 | 14.8 | 74.0 | 598 | 6 | ABK16211 | Human lun |
| C 291 | 14.8 | 74.0 | 598 | 9 | AB954474 | Human lun |
| C 292 | 14.8 | 74.0 | 610 | 3 | ABH31084 | Human col |
| C 293 | 14.8 | 74.0 | 616 | 6 | ABL93019 | Rat metas |
| C 294 | 14.8 | 74.0 | 620 | 3 | AAAL6138 | Human col |
| C 295 | 14.8 | 74.0 | 620 | 3 | AAAC03957 | Human col |
| C 296 | 14.8 | 74.0 | 624 | 6 | ABK33069 | DNA encod |
| C 297 | 14.8 | 74.0 | 638 | 4 | AAAS24470 | Human ova |
| C 298 | 14.8 | 74.0 | 638 | 5 | AAH83074 | Human ova |
| C 299 | 14.8 | 74.0 | 638 | 2 | AAH97692 | Extended |
| C 300 | 14.8 | 74.0 | 648 | 7 | AB236503 | Human GEN |

ALIGNMENTS

| | | | | | | |
|--|---------------|---------|---------|--|--|--|
| RESULT 1 | | | | | | |
| AAI99682_25 | | | | | | |
| Continuation (26 of 45) of AAI99682 from base 2500001 (Mycobacterium tuberculosis strain | | | | | | |
| WP Sequence split into 45 fragments | | | | | | |
| WP | Fragment Name | Begin | End | | | |
| WP | AAI99682_00 | 1 | 110000 | | | |
| WP | AAI99682_01 | 100001 | 210000 | | | |
| WP | AAI99682_02 | 200001 | 310000 | | | |
| WP | AAI99682_03 | 300001 | 410000 | | | |
| WP | AAI99682_04 | 400001 | 510000 | | | |
| WP | AAI99682_05 | 500001 | 610000 | | | |
| WP | AAI99682_06 | 600001 | 710000 | | | |
| WP | AAI99682_07 | 700001 | 810000 | | | |
| WP | AAI99682_08 | 800001 | 910000 | | | |
| WP | AAI99682_09 | 900001 | 1010000 | | | |
| WP | AAI99682_10 | 1000001 | 1110000 | | | |
| WP | AAI99682_11 | 1100001 | 1210000 | | | |
| WP | AAI99682_12 | 1200001 | 1310000 | | | |
| WP | AAI99682_13 | 1300001 | 1410000 | | | |
| WP | AAI99682_14 | 1400001 | 1510000 | | | |
| WP | AAI99682_15 | 1500001 | 1610000 | | | |
| WP | AAI99682_16 | 1600001 | 1710000 | | | |
| WP | AAI99682_17 | 1700001 | 1810000 | | | |

| | | | | | | |
|----|-------------|---------|---------|--|--|--|
| WP | AAI99682_05 | 500001 | 610000 | | | |
| WP | AAI99682_06 | 600001 | 710000 | | | |
| WP | AAI99682_07 | 700001 | 810000 | | | |
| WP | AAI99682_08 | 800001 | 910000 | | | |
| WP | AAI99682_09 | 900001 | 1010000 | | | |
| WP | AAI99682_10 | 1000001 | 1110000 | | | |
| WP | AAI99682_11 | 1100001 | 1210000 | | | |
| WP | AAI99682_12 | 1200001 | 1310000 | | | |
| WP | AAI99682_13 | 1300001 | 1410000 | | | |
| WP | AAI99682_14 | 1400001 | 1510000 | | | |
| WP | AAI99682_15 | 1500001 | 1610000 | | | |
| WP | AAI99682_16 | 1600001 | 1710000 | | | |
| WP | AAI99682_17 | 1700001 | 1810000 | | | |
| WP | AAI99682_18 | 1800001 | 1910000 | | | |
| WP | AAI99682_19 | 1900001 | 2010000 | | | |
| WP | AAI99682_20 | 2000001 | 2110000 | | | |
| WP | AAI99682_21 | 2100001 | 2210000 | | | |
| WP | AAI99682_22 | 2200001 | 2310000 | | | |
| WP | AAI99682_23 | 2300001 | 2410000 | | | |
| WP | AAI99682_24 | 2400001 | 2510000 | | | |
| WP | AAI99682_25 | 2500001 | 2610000 | | | |
| WP | AAI99682_26 | 2600001 | 2710000 | | | |
| WP | AAI99682_27 | 2700001 | 2810000 | | | |
| WP | AAI99682_28 | 2800001 | 2910000 | | | |
| WP | AAI99682_29 | 2900001 | 3010000 | | | |
| WP | AAI99682_30 | 3000001 | 3110000 | | | |
| WP | AAI99682_31 | 3100001 | 3210000 | | | |
| WP | AAI99682_32 | 3200001 | 3310000 | | | |
| WP | AAI99682_33 | 3300001 | 3410000 | | | |
| WP | AAI99682_34 | 3400001 | 3510000 | | | |
| WP | AAI99682_35 | 3500001 | 3610000 | | | |
| WP | AAI99682_36 | 3600001 | 3710000 | | | |
| WP | AAI99682_37 | 3700001 | 3810000 | | | |
| WP | AAI99682_38 | 3800001 | 3910000 | | | |
| WP | AAI99682_39 | 3900001 | 4010000 | | | |
| WP | AAI99682_40 | 4000001 | 4110000 | | | |
| WP | AAI99682_41 | 4100001 | 4210000 | | | |
| WP | AAI99682_42 | 4200001 | 4310000 | | | |
| WP | AAI99682_43 | 4300001 | 4410000 | | | |
| WP | AAI99682_44 | 4400001 | 4411529 | | | |

Query Match 95.0%; Score 19; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GAGGCGCCGAGCCCAAA 20
Db 104135 GAGGCGCCGAGCCCAAA 104153

RESULT 2
AAI99682_26
Continuation (27 of 45) of AAI99682 from base 2600001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments

| | | | | | | |
|----|---------------|---------|---------|--|--|--|
| WP | Fragment Name | Begin | End | | | |
| WP | AAI99682_00 | 1 | 110000 | | | |
| WP | AAI99682_01 | 100001 | 210000 | | | |
| WP | AAI99682_02 | 200001 | 310000 | | | |
| WP | AAI99682_03 | 300001 | 410000 | | | |
| WP | AAI99682_04 | 400001 | 510000 | | | |
| WP | AAI99682_05 | 500001 | 610000 | | | |
| WP | AAI99682_06 | 600001 | 710000 | | | |
| WP | AAI99682_07 | 700001 | 810000 | | | |
| WP | AAI99682_08 | 800001 | 910000 | | | |
| WP | AAI99682_09 | 900001 | 1010000 | | | |
| WP | AAI99682_10 | 1000001 | 1110000 | | | |
| WP | AAI99682_11 | 1100001 | 1210000 | | | |
| WP | AAI99682_12 | 1200001 | 1310000 | | | |
| WP | AAI99682_13 | 1300001 | 1410000 | | | |
| WP | AAI99682_14 | 1400001 | 1510000 | | | |
| WP | AAI99682_15 | 1500001 | 1610000 | | | |
| WP | AAI99682_16 | 1600001 | 1710000 | | | |
| WP | AAI99682_17 | 1700001 | 1810000 | | | |

Db 131 CGAGGGCCCAATCCCAACAA 112
 RESULT 5
 ABQ75352/c
 ID A5Q75352 standard; cDNA; 993 BP.
 XX
 DE A5Q75352 standard; cDNA; 993 BP.
 XX
 AC A5Q75352;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE Human lung specific nucleic acid sequence SEQ ID NO:91.
 XX
 KW Human; lung; lung specific nucleic acid; LSNA; lung specific protein; LSP;
 XX LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
 KW squamous cell carcinoma; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WC200264788-A2.
 XX
 PD 22-AUG-2002.
 XX
 XX 20-NOV-2001; 2001WO-US045090.
 XX
 PF 20-NOV-2000; 2000US-0252054P.
 XX
 PR 20-NOV-2000; 2000US-0252054P.
 XX
 XX (DIAD-) DIADEXUS INC.
 PA
 XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;
 PI WPI; 2002-657601/70.
 XX
 DR New lung specific nucleic acid useful in gene therapy or as vaccines for
 XX treating lung cancer (e.g. squamous cell carcinoma) or non-cancerous lung
 PT diseases, as well as for diagnosing, monitoring or staging these
 PT diseases.
 XX
 PS Claim 1; Page 206-207; 282pp; English.
 XX
 CC The present invention describes an isolated lung specific nucleic acid
 CC (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
 CC sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
 CC (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
 CC given in ABQ75362 to ABQ75376; (c) selectively hybridises to (a) or (b);
 CC or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
 CC protein (LSP) sequences have cytostatic activity and can be used in gene
 CC therapy and vaccines. LSNA and LSPs are useful for diagnosing and
 CC monitoring the presence and metastases of lung cancer in a patient. An
 CC antibody that specifically binds to an LSP can be used for determining
 CC the presence of an LSP in a sample, as well as for treating a patient
 CC with lung cancer, particularly by inducing an immune response against the
 CC lung cancer cell expressing the LSNA or LSPs. In particular, these LSNA
 CC and LSPs are useful for identifying, diagnosing, monitoring, staging,
 CC imaging and treating lung cancer (e.g. squamous cell carcinoma) and non-
 CC cancerous disease states in lung
 XX
 SQ Sequence 993 BP; 210 A; 225 C; 312 G; 246 T; 0 U; 0 Other;
 Query Match 84.0%; Score 16.8; DB 6; Length 993;
 Best Local Similarity 90.0%; Pred. No. 5.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CGAGGGCCCAATCCCAACAA 20
 |||||
 Db 485 CGCGGGCCCAATCCCAACAA 466
 RESULT 6
 AAD39099/c
 ID AAD39099 standard; cDNA; 1002 BP.
 XX
 AC AAD39099;
 XX
 DT 19-JUN-2003 (first entry)
 XX


```

XX Treponema pallidum.
XX OS WO9859034-A2.
XX PN
XX PD
XX PP 30-DEC-1998.
XX PP 23-JUN-1998; 98WO-US013041.
XX PR 24-JUN-1997; 97US-0050667P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Fraser CM;
XX PI WPI; 1999-081273/07.
XX DR
XX PT New isolated Treponema pallidum nucleic acids - used to develop products
XX PT for the detection, diagnosis, characterisation, prevention and therapy of
XX PT T. pallidum infections, particularly syphilis.
XX PS Claim 1; Page 736-747; 1150pp; English.
XX CC AAX20500-21243 represent polynucleotide sequences from the genome of
XX CC Treponema pallidum. The sequences can be used for detection, diagnosis,
XX CC characterisation, prevention and therapy for T. pallidum infections,
XX CC particularly syphilis. They can also be used for detecting diseases
XX CC related to Borrelia infections in animals, and for the production of
XX CC biosynthetic products such as enzymes
XX CC
XX SQ Sequence 19186 BP; 4705 A; 5450 C; 4881 G; 4131 T; 0 U; 19 Other;

Query Match 84.0%; Score 16.8; DB 2; Length 19196;
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCGAGCCCAAA 20
DB 9003 CGAGGCGCCGAGCCCAAA 9022

RESULT 10
ADC35539/C
ID ADC35539 standard; DNA; 21501 BP.
XX AC
XX AC ADC35539;
XX DT 18-DEC-2003 (first entry)
XX DE Human CD81/TAPA-1 partial genomic sequence.
XX KW Antisense; ds; human; CD81; TAPA-1; tetraspanin; viral infection;
XX KW cocaine addiction; autoimmune disorder; antiinflammatory; antibacterial;
XX KW virucide; antiparasitic; inflammatory disorder; parasitic infection;
XX KW bacterial infection.
XX OS Homo sapiens.
XX PN US2003113914-A1.
XX PD 19-JUN-2003.
XX PF 10-DEC-2001; 2001US-00006430.
XX PR 10-DEC-2001; 2001US-00006430.
XX PA (ISIS-) ISIS PHARM INC.
XX PI Graham MJ, Dobie K;
XX PI WPI; 2003-810907/76.
XX PT Novel compound hybridizing with nucleic acid molecule encoding CD81 and

```

```

PT inhibiting the expression of CD81, useful for treating infections and
PT disease associated with expression of CD81 such as inflammation disorder.
XX PS Example 15; SEQ ID NO 11; 55pp; English.
XX CC The invention relates to a compound (antisense oligonucleotide)
XX CC hybridising with the eighth nucleobase portion of an active site on a
XX CC nucleic acid molecule encoding CD81 (also known as TAPA-1, a tetraspanin)
XX CC and inhibiting the expression of CD81. Also included is a composition
XX CC comprising the antisense oligonucleotide and a carrier or a diluent. The
XX CC antisense oligonucleotide is useful for inhibiting the expression of CD81
XX CC in cells or tissues. The antisense oligonucleotide is also useful for
XX CC treating infections preferably viral, bacterial and parasitic and
XX CC diseases such as inflammatory disorders and autoimmune disorders. The
XX CC disease or condition is characterised by chemical dependency (e.g.
XX CC cocaine addiction). The present sequence is a Human CD81/TAPA-1 partial
XX CC genomic sequence.
XX SQ Sequence 21501 BP; 3604 A; 6408 C; 7119 G; 4370 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 9; Length 21501;
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCGAGCCCAAA 20
DB 7509 CRAGGCGCCGAGCCCAAA 7490

RESULT 11
ADD00956
ID ADD00956 standard; DNA; 28000 BP.
XX AC
XX AC ADD00956;
XX DT 01-JAN-2004 (first entry)
XX DE Human Jagged 2 encoding DNA SEQ ID NO:11.
XX KW apoptosis; Jagged 2 inhibitor; cytostatic; hyperproliferative disorder;
XX KW human; ds; gene; chromosome 14.
XX OS Homo sapiens.
XX PN WO2003077848-A2.
XX PD 25-SEP-2003.
XX PF 10-MAR-2003; 2003WO-US007340.
XX PR 12-MAR-2002; 2002US-00096399.
XX PA (ISIS-) ISIS PHARM INC.
XX PI Koller E, Shapard PJ;
XX PI WPI; 2003-756943/71.
XX PD P-PSDB; ADD01037.
XX PT Inducing apoptosis in a cell or animal for treating a subject having a
XX PT condition associated with insufficient apoptosis by administering to a
XX PT cell or animal a Jagged 2 inhibitor to reduce Jagged 2 levels or
XX PT activity.
XX PS Example 15; SEQ ID NO 11; 148pp; English.
XX CC The present invention describes a method for inducing apoptosis in a cell
XX CC or animal comprising administering to a cell or animal a Jagged 2
XX CC inhibitor to reduce Jagged 2 levels or activity. Also described: (1)
XX CC treating a subject having a disease or condition associated with
XX CC insufficient apoptosis by administration of a Jagged 2 inhibitor; (2) a
XX CC pharmaceutical composition comprising a Jagged 2 inhibitor and another
XX CC active ingredient for inducing apoptosis; and (3) a kit comprising a

```

CC Jagged 2 inhibitor and instructions for using the Jagged 2 inhibitor in
 CC the induction of apoptosis. The Jagged 2 inhibitor has cytostatic
 CC activity. The method can be used for inducing apoptosis in a cell or
 CC animal for treating a subject having a disease or condition associated
 CC with insufficient apoptosis, e.g., hyperproliferative disorder. The
 CC present sequence encodes the human Jagged 2 protein, which is used in an
 CC example from the present invention. Human Jagged 2 is located on
 CC chromosome 14.
 XX
 SQ Sequence 28000 BP; 4121 A; 8288 C; 10292 G; 5299 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 9; Length 28000;
 Best Local Similarity 90.0%; Pred. No. 5.3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGGCGCCGAGCCCAACAA 20
 Db 26057 CCAGCGCGCCGAGCCCAACAA 26076

RESULT 12
 ABQ75121/c
 ID ABQ75121 standard; DNA; 37113 BP.

XX AC ABQ75121;

XX DT 04-NOV-2002 (first entry)

XX DE Human CD81 gene SEQ ID NO:11.

XX KW Human; p15 region; chromosome 11; tumour growth; infertility; cytostatic;
 XX KW hepatitis C virus infection; antiinfertility; gene therapy; CD 81;
 XX KW cluster of differentiation antigen 81; gene; ds.

XX OS Homo sapiens.

XX FH Location/Qualifiers

FT CDS 10471..29787
 FT /tag= a
 FT /product= "cluster of differentiation antigen 81 (CD 81)
 FT protein"

FT /note= "contains introns"

FT exon 10471..10536

FT /tag= b

FT /number= 1

FT intron 10537..23332

FT /tag= c

FT exon 23333..23446

FT /tag= d

FT intron 23447..27014

FT /tag= e

FT exon 27015..27113

FT /tag= f

FT intron 27114..27892

FT /tag= e

FT exon 27893..27964

FT /tag= f

FT intron 27963..28333

FT /tag= e

FT exon 28334..28441

FT /tag= f

FT intron 28442..28789

FT /tag= e

FT exon 28790..28891

FT /tag= f
 FT /number= 6
 FT intron 28892..29548
 FT /tag= e
 FT /number= 6
 FT exon 29549..29635
 FT /tag= f
 FT /number= 7
 FT intron 29636..29724
 FT /tag= e
 FT /number= 7
 FT exon 29725..29784
 FT /tag= f
 FT /number= 8

XX WO200261085-A2.

XX PD 08-AUG-2002.

XX 31-OCT-2001; 2001WO-US045381.

XX 31-OCT-2000; 2000US-0244705P.

XX (RYAN/) RYAN J W.

XX PI Ryan JW;

XX DR WPI; 2002-619251/66.

XX P-PSDB; ABP52845.

XX New genes obtainable from the p15 region of human chromosome 11 (e.g.
 XX human achaete-scute homolog 2), useful in gene therapy, particularly for
 XX preventing or treating tumor growth, infertility or hepatitis C virus
 XX infection.

XX Claim 1; Fig 2; 94pp; English.

XX The present invention describes an isolated genomic polynucleotide (I),
 XX which is obtainable from the p15 region of human chromosome 11.
 XX Specifically described are the human achaete-scute homologue 2 (HASH2),
 XX SMS3, tumour suppressing sub-transferable candidate 6 (TSSC6), ribosomal
 XX protein L26 (RIBO26), cluster of differentiation antigen 81 (CD 81) and
 XX tumour suppressing sub-transferable candidate 4 (TSSC4) genes as given in
 XX ABQ75117 to ABQ75122, encoding the proteins given in ABP52841 to
 XX ABP52846. Also described is an isolated polynucleotide obtainable from
 XX the p15 region of chromosome 11 having the sequence, which comprises any
 XX of two nucleotide sequences given in ABQ75123 and ABQ75124. (I) has
 XX cytostatic and antiinfertility activities and can be used in gene
 XX therapy. The HASH2, SMS3, TSSC6, RIBO26, CD 81 or TSSC4 polynucleotide
 XX sequences can be used for preventing, treating or ameliorating a medical
 XX condition, or for manufacturing a gene therapy for the prevention,
 XX treatment or amelioration of such medical conditions. An antisense
 XX oligonucleotide or mimetic of (I) can also be useful for manufacturing a
 XX medicament for the prevention, treatment or amelioration of these medical
 XX conditions. These conditions include tumour growth, infertility, or
 XX hepatitis C virus infection

SQ Sequence 37113 BP; 6829 A; 10985 C; 11249 G; 8050 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 6; Length 37113;

Best Local Similarity 90.0%; Pred. No. 5.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGGCGCCGAGCCCAACAA 20

Db 17233 CAAGGGCGCCGAGCCCAACAA 17214

RESULT 13

AAC10189

ID AAC10189 standard; cDNA; 354 BP.

XX AC AAC10189;

```

XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 14264.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 14264; 71bp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 354 BP; 58 A; 155 C; 85 G; 53 T; 0 U; 3 Other;
XX
XX Query Match 82.0%; Score 16.4; DB 3; Length 354;
XX Best Local Similarity 94.4%; Pred. No. 8.2e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 CGAGGCGCCCGCCCGCCAC 18
XX 306 CGAGGCGCCCGCCCGCCAC 323
XX
XX RESULT 14
XX AAC59241.
XX ID AAC59241 standard; cDNA; 3145 BP.
XX
XX AAC59241;
XX
XX 29-JAN-2001 (first entry)
XX
XX Human secreted protein cDNA sequence #27.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; ss.
XX Homo sapiens.

```

```

XX
XX WO200055199-A1.
XX
XX 21-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US006014.
XX
XX 12-MAR-1999; 99US-0124095P.
XX 11-JUN-1999; 99US-0138598P.
XX 03-DEC-1999; 99US-0168665P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-572359/53.
XX P-FSDB; AAB27820.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; Page 355-356; 433pp; English.
XX
XX The invention relate to the isolation of genes AAC59215-CS9261 encoding
XX 47 human secreted proteins AAB27794-B27840. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion for increasing the stability of the fusion
XX protein as compared to the human protein only. The genes and proteins are
XX useful for preventing, ameliorating or treating medical conditions, e.g.
XX by protein or gene therapy. The genes are isolated from a range of human
XX tissues disclosed in the specification. The nucleic acids, proteins,
XX antibodies and (ant)agonists are useful in the diagnosis, treatment and
XX prevention of: (a) cancer, e.g. breast and ovarian cancer, and other
XX cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal
XX tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders
XX such as myocardial ischaemias; (d) wound healing; (e) neurological
XX diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases
XX such as viral, bacterial, fungal and parasitic infections
XX
XX Sequence 3145 BP; 806 A; 1006 C; 745 G; 581 T; 0 U; 7 Other;
XX
XX Query Match 82.0%; Score 16.4; DB 3; Length 3145;
XX Best Local Similarity 94.4%; Pred. No. 8e-02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 CGAGGCGCCCGCCCGCCAC 18
XX 37 CGAGGCGCCCGCCCGCCAC 54
XX
XX RESULT 15
XX AAS99902
XX ID AAS99902 standard; cDNA; 3287 BP.
XX
XX AAS99902;
XX
XX 12-MAR-2002 (first entry)
XX
XX Polynucleotide encoding human cytoskeleton-associated protein #13.
XX
XX Human; cytoskeleton-associated protein; CYSKP; autoimmune disorder; ss;
XX cell proliferative disorder; inflammatory disorder; prion disease;
XX vesicle trafficking disorder; gastrointestinal disorder; muscle disorder;
XX neurological disorder; cell motility disorder; reproductive disorder;
XX spinal cord disease; central nervous system disorder; mental disorder;
XX gene therapy; cancer.
XX Homo sapiens.
XX
XX WO200185942-A2.

```

XX 15-NOV-2001.
 XX PD
 XX PF
 XX 03-MAY-2001; 2001WO-US014355.
 XX
 XX 05-MAY-2000; 2000US-0201960P.
 XX PR
 XX 08-MAY-2000; 2000US-0202729P.
 XX PR
 XX 05-JUN-2000; 2000US-0209705P.
 XX PR
 XX 07-JUN-2000; 2000US-0210149P.
 XX PR
 XX 21-JUN-2000; 2000US-0213215P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX PA
 XX Yue H, Tang YT, Au-Young J, Lu DAM, Baughn MR, Hillman JL;
 XX PI Azimzai Y, Lal P, Yao MG, Bandman O, Burford N, Batra S, Kearney L;
 XX PI Policky JL;
 XX DR WPI; 2002-062248/08.
 XX DR P-PSDB; AAU74342.
 XX
 XX New cytoskeleton-associated proteins and polynucleotides, useful for
 XX PT diagnosing, preventing and treating cell proliferative, autoimmune,
 XX PT inflammatory, neurological, cell motility, reproductive and muscle
 XX PT disorders.
 XX PS
 XX Claim 5; Page 178; 194pp; English.
 XX
 XX The invention relates to human cytoskeleton-associated polypeptides
 XX (CYSKP) and their associated polynucleotide sequences. The sequences are
 XX useful in the treatment of disorders associated with overexpression or
 XX underexpression of CYSKP in a patient. The disorders include cell
 XX proliferative disorders (such as cancer, actinic keratosis,
 XX CC arteriosclerosis, cirrhosis, hepatitis and psoriasis),
 XX CC autoimmune/inflammatory disorders (such as, asthma, atherosclerosis,
 XX CC osteoporosis, Crohn's disease, rheumatoid arthritis, diabetes mellitus
 XX CC and anaemia), vesicle trafficking disorders (such as
 XX CC hypercholesterolaemia, diabetes insipidus, grave's disease and goitre),
 XX CC gastrointestinal disorders, prion diseases, neurological disorders (such
 XX CC as epilepsy, stroke, cerebral neoplasms, Alzheimer's disease,
 XX CC Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis
 XX CC and other motor neuron disorders), cell motility disorders, reproductive
 XX CC disorders (such as endometriosis and polycystic ovary syndrome), muscle
 XX CC disorders (such as myocardiitis, migraine, hypertension, hypoglycaemia,
 XX CC myocardial infarction, epilepsy and muscular dystrophy), spinal cord
 XX CC diseases, central nervous system disorders (such as Down syndrome and
 XX CC cerebral palsy) and mental disorders (such as anxiety and schizophrenia).
 XX CC Sequences AAS99890-AAS99923 represent cDNA molecules encoding human CYSKP
 XX CC of the invention
 XX SQ Sequence 3287 BP; 795 A; 1095 C; 788 G; 609 T; 0 U; 0 Other;
 Query Match 82.0%; Score 16.4; DB 6; Length 3287;
 Best Local Similarity 94.4%; Pred. No. 8e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 CGAGGCGCCCGAGCCCGAC 18
 DB 220 CGAGGCGCCCGAGCCCGAC 237
 RESULT 16
 ABN59766
 ID ABN59766 standard; cDNA; 3345 BP.
 XX AC ABN59766;
 XX AC
 XX 28-JUN-2002 (first entry)
 XX DT
 XX DE Novel human coding sequence SEQ ID NO: 177.
 XX
 XX Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
 XX KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
 XX KW neuroprotective; antiparkinsonian; protein therapy; EST;

XX expressed sequence tag; gene; ss.
 XX OS Homo sapiens.
 XX PN WO200222660-A2.
 XX XX
 XX PD 21-MAR-2002.
 XX XX
 XX PF 10-SEP-2001; 2001WO-US026015.
 XX XX
 XX PR 11-SEP-2000; 2000US-00659671.
 XX XX
 XX (HYSE-) HYSEQ INC.
 XX PA
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 XX PI Xue AJ, Yang Y, Wehrman T, Dmanac RT;
 XX DR WPI; 2002-292408/33.
 XX DR P-PSDB; ABB97353.
 XX
 XX An isolated polynucleotide for treating diseases associated with its
 XX PT encoded polypeptide such as cancer and multiple sclerosis.
 XX PT
 XX Claim 1; SEQ ID NO 177; 509pp; English.
 XX PS
 XX The present invention provides the protein and coding sequences of 444
 XX CC novel human proteins. These were isolated from expressed sequences tags
 XX CC (ESTs). They can be used to stimulate cell growth, to regulate
 XX CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 XX CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 XX CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 XX CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
 XX CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
 XX CC rheumatoid arthritis, and to treat nervous system disorders e.g.
 XX CC Parkinson's disease. The present sequence is a coding sequence of the
 XX CC invention
 XX SQ Sequence 3345 BP; 810 A; 1120 C; 792 G; 623 T; 0 U; 0 Other;
 Query Match 82.0%; Score 16.4; DB 6; Length 3345;
 Best Local Similarity 94.4%; Pred. No. 8e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 CGAGGCGCCCGAGCCCGAC 18
 DB 274 CGAGGCGCCCGAGCCCGAC 291
 RESULT 17
 ABZ24596
 ID ABZ24596 standard; DNA; 19951 BP.
 XX AC ABZ24596;
 XX AC
 XX 31-MAR-2003 (first entry)
 XX DT
 XX DE Human ovary-specific O1-236 gene.
 XX XX
 XX Ovary; O1-236; human; nucleoplasmin 2; Npm2; contraceptive;
 XX KW antinfertility; cytostatic; gene therapy; gene; ds.
 XX OS Homo sapiens.
 XX XX
 XX Key Location/Qualifiers
 XX FT exon 2593..2640
 XX FT /tag= b
 XX FT intron 2641..2769
 XX FT /tag= c
 XX FT exon 2770..2855
 XX FT /tag= d
 XX FT intron 2856..2980
 XX FT /tag= e
 XX FT exon 2981..3106

```
FT intron /*tag= f
FT 3107. .10463
FT /*tag= g
FT 10464. .10557
FT /*tag= h
FT 10558. .11422
FT /*tag= i
FT 11443. .11608
FT /*tag= j
FT 11609. .11877
FT /*tag= k
FT 11833. .11877
FT /*tag= l
FT 11878. .13925
FT /*tag= m
FT 13826. .13859
FT /*tag= n
FT 13860. .13970
FT /*tag= o
FT 13971. .14015
FT /*tag= p
FT 14015
FT CDS /*tag= a
FT /*product= "Human NPM2"
FT /*note= "contains introns"
XX
XX WO200298314-A2.
XX
XX 07-NOV-2002.
XX
XX 26-APR-2002; 2002WO-US013245.
XX
XX 27-APR-2001; 2001US-00844864.
XX (BAY) BAYLOR COLLEGE MEDICINE.
XX (AMHP) WYETH.
XX
XX Matzuk MM, Wang P, Bai Y, Wu X;
XX
XX WPI; 2003-167110/16.
XX P-PSDB; ABP58238.
XX
XX New ovary-specific-genes comprising OI-180 or OI-236, useful for
XX decreasing conception or enhancing fertility, or for the preparation of a
XX composition for treating e.g. cancer.
XX
XX Disclosure; Page 119-130; 141pp; English.
XX
XX The present sequence is that of the human ovary-specific OI-236 gene. OI-
XX 236 is the mammalian orthologue of Xenopus laevis nucleoplasmin, and the
XX OI-236 gene has been named Npm2. Experiments in mice have shown that loss
XX of Npm2 results in female infertility and subfertility. The invention
XX provides ovary-specific and oocyte-specific murine and human OI-180, OI-
XX 184 and OI-236 polynucleotides and polypeptides. These genes and their
XX protein products appear to relate to various cell proliferative or
XX degenerative disorders, especially those involving ovarian tumours, such
XX as germ line tumours and granulosa cell tumours, or infertility, such as
XX premature ovarian failure. The invention provides a method for detection
XX of a cell proliferative or degenerative disorder of the ovary, which is
XX associated with the expression of OI-180, OI-184 or OI-236. It also
XX provides a method for treating such disorders by using an agent which
XX suppresses or enhances the respective activities of OI-180, OI-184 or OI-
XX 236, and a method of screening for compounds that interact and/or
XX modulate the expression or activity of the ovary-specific genes. These
XX compounds are possible contraceptive or fertility enhancing agents. The
XX modulator is preferably a polypeptide, small molecule or polynucleotide
XX sequence
XX
XX Sequence 19951 BP; 4619 A; 5201 C; 4927 G; 5204 T; 0 U; 0 Other;
XX
XX Query Match 82.0%; Score 16.4; DB 7; Length 19951;
XX Best Local Similarity 94.4%; Pred. No. 7.9e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX Query Match 80.0%; Score 16; DB 6; Length 3417;
XX Best Local Similarity 100.0%; Pred. No. 1.2e+03;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Query 5 GCGCCCGAGCCCAAA 20
XX |||||
XX Db 657 GCGCCCGAGCCCAAA 642
XX
XX RESULT 19
XX AB235700/c
XX ID AB235700 standard; DNA; 3417 BP.
XX
XX AC AB235700;
```

```
QY 2 GAGCGCCCGAGCCCAAA 19
Db 14784 GAGCGCCCGAGCCCAAA 14801
```

```
RESULT 18
ABV78124/c
ID ABV78124 standard; DNA; 3417 BP.
XX
XX AC ABV78124;
XX
XX DT 15-NOV-2002 (first entry)
XX
XX DE Human Tiel DNA SEQ ID NO 8.
XX
XX KW RNA inhibition; dsRNA; gene expression inhibitor; oncogene; cytostatic;
XX virucide; protozoacide; gene; ds.
XX
XX OS Homo sapiens.
XX
XX FN WO200255693-A2.
XX
XX PD 18-JUL-2002.
XX
XX PF 09-JAN-2002; 2002WO-EF000152.
XX
XX PR 09-JAN-2001; 2001DE-01000586.
XX 26-OCT-2001; 2001DE-01055280.
XX PR 29-NOV-2001; 2001DE-01058411.
XX PR 07-DEC-2001; 2001DE-01060151.
XX
XX PA (RIBO-) RIBOPHARMA AG.
XX
XX PI Kreutzer R, Lämmer S, Rost S, Radwiger P;
XX WPI; 2002-590671/63.
XX
XX DR
XX
XX PT Inhibiting expression of target gene, useful e.g. for inhibiting
XX oncogenes, by administering double-stranded RNA complementary to the
XX target and having an overhang.
XX
XX PS Claim 10; Page 116-117; 203pp; German.
XX
XX CC The invention relates to inhibiting expression of a target gene (I) in a
XX cell by introducing an inhibitory RNA (dsRNA1) having a double-stranded
XX structure of at most 49 consecutive bases. At least part of one strand
XX (as1) of dsRNA1 is complementary to (I) and at least one end of dsRNA1
XX has an overhang of 1-4 nucleotides. The method is used to inhibit the
XX expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.
XX in humans, also genes in plasmidium or in viruses or viroids that are
XX pathogenic for humans, animals or plants. Introducing an overhang into
XX dsRNA greatly increases effectiveness for inhibiting gene expression,
XX both in vivo and in vitro and also increases stability and thus the
XX effective concentration inside the cell. The present sequence is that of
XX a gene related to the invention
XX
XX SQ Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;
```

```
Query Match 80.0%; Score 16; DB 6; Length 3417;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 GCGCCCGAGCCCAAA 20
|||
Db 657 GCGCCCGAGCCCAAA 642
```

```
RESULT 19
AB235700/c
ID AB235700 standard; DNA; 3417 BP.
XX
XX AC AB235700;
```

XX DT 07-FEB-2003 (first entry)
 XX XX Human Tiel encoding polynucleotide SEQ ID NO 8.
 XX DE Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;
 KW Protozoacide; gene expression; antisense; tumour; infection; Plasmodium;
 KW virus; viroid; anti-GFP; human; HIV; human immunodeficiency virus;
 KW Hepatitis C virus; human papilloma virus; gene; ds.
 XX OS Homo sapiens.
 XX XX DE10100588-A1.
 XX PN 18-JUL-2002.
 XX PD 09-JAN-2001; 2001DE-01000588.
 XX PF 09-JAN-2001; 2001DE-01000588.
 XX PR 09-JAN-2001; 2001DE-01000588.
 XX PS (RIBO-) RIBOPHARMA AG.
 XX PA Kreutzer R, Limmer S, Rost S, Hadwiger P;
 XX PI WPI; 2002-683450/74.
 XX DR Inhibiting expression of target genes, useful e.g. for treating tumors,
 PT by introducing into cells two double-stranded RNAs that are complementary
 PT to the target.
 XX PS Claim 13; Page 12-13; 100pp; German.
 XX CC The invention relates to inhibiting expression of a target gene in a cell
 CC by introducing at least two oligoribonucleotides (dsRNAI and II), both
 CC with a double-stranded (ds) structure of at most 49 sequential nucleotide
 CC pairs. At least part of one strand (S1, S2) of the ds structures in each
 CC of dsRNAI and II are complementary to regions in the target gene. The
 CC method uses antisense inhibition of gene expression using double stranded
 CC RNA inhibition (RNAI). The method is particularly used to treat tumours
 CC or infections, especially by Plasmodium or viruses/viroids (pathogenic on
 CC humans, animals or plants). The method provides more effective inhibition
 CC of expression than known methods using a single dsRNA, even at very low
 CC concentrations. When dsRNA has at least one unpaired nucleotide at the
 CC end, stability (and thus effective concentration in the cell) is improved
 CC and efficiency can be increased further by pretreating the cells with
 CC interferon. The present sequence is that of a target DNA of the invention
 XX
 XX SQ Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;
 Query Match 80.0%; Score 16; DB 6; Length 3417;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 GCGCCCGAGCCCAACAA 20
 DB 657 GCGCCCGAGCCCAACAA 642
 RESULT 20
 ABX09943/C
 ID ABX09943 standard; DNA; 3417 BP.
 XX AC ABX09943;
 XX XX 23-JAN-2003 (first entry)
 XX DT Human Tiel DNA fragment SEQ ID 8.
 XX DE Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;
 KW prion; inhibition; human; ds.
 XX OS Homo sapiens.
 XX XX

PN DE10100587-Cl.
 XX 21-NOV-2002.
 XX PF 09-JAN-2001; 2001DE-01000587.
 XX PR 09-JAN-2001; 2001DE-01000587.
 XX PA (RIBO-) RIBOPHARMA AG.
 XX PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
 XX XX WPI; 2002-742209/81.
 XX DR Inhibiting expression of target genes, e.g. oncogenes, in cells, by
 PT introduction of complementary double-stranded oligoribonucleotide, after
 PT treating the cell with interferon.
 XX PS Disclosure; Page 17-18; 98pp; German.
 XX CC This invention describes a novel method for inhibiting expression of a
 CC target gene by introducing into the cell that contains the target gene at
 CC least one oligoribonucleotide (dsRNAI) that has a double-stranded (ds)
 CC structure of not more than 49 consecutive nucleotides (nt), where at
 CC least a segment of one strand of the ds structure is complementary with
 CC the target gene and the cells are treated with interferon before
 CC introduction of dsRNAI. The method is used to inhibit expression of
 CC target genes, particularly oncogenes, cytokine genes, Id (not defined)
 CC protein genes; developmental or prion genes, or genes expressed in
 CC pathogenic organisms (particularly plasmodia) or in viruses or viroids
 CC (pathogenic in humans, animals or plants). Treating the cells with
 CC interferon greatly increases the extent to which dsRNA can inhibit
 CC expression of the target genes, and the effect is even greater when dsRNA
 CC are modified to increase their stability. ABX09936-ABX10075 represent
 CC gene fragments used to illustrate the method of the invention
 XX
 XX SQ Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;
 Query Match 80.0%; Score 16; DB 6; Length 3417;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 GCGCCCGAGCCCAACAA 20
 DB 657 GCGCCCGAGCCCAACAA 642
 RESULT 21
 ABL91665/C
 ID ABL91665 standard; DNA; 3417 BP.
 XX AC ABL91665;
 XX XX 28-MAY-2002 (first entry)
 XX DT Human polynucleotide SEQ ID NO 8.
 XX DE Human; HIV; HCV; gene expression; oligoribonucleotide; tumour; pathogen;
 KW Plasmodium; virus; viroid; cytokine; prion; antisense oligonucleotide;
 KW cytostatic; virucide; protozoacide; antibacterial; ds.
 XX OS Homo sapiens.
 XX XX DE10100586-Cl.
 XX PN 11-APR-2002.
 XX PD 09-JAN-2001; 2001DE-01000586.
 XX PF 09-JAN-2001; 2001DE-01000586.
 XX PR (RIBO-) RIBOPHARMA AG.
 XX PA

```
PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
XX WPI; 2002-270454/32.
XX
XX Inhibiting gene expression in cells, useful for e.g. treating tumors, by
XX introducing double-stranded complementary oligonucleotides having unpaired
XX terminal bases.
XX
XX Claim 13; Page 12-13; 104pp; German.
XX
XX The invention relates to a method for inhibiting expression of a target
XX gene (ASL91658-AB191797) in a cell by introducing at least one
XX oligonucleotide that has a double-stranded structure consisting of at
XX most 49 sequential nucleotide pairs, with at least part of one strand
XX complementary with the target gene and has at least one end a single-
XX stranded segment of 1-4 nt. The method provides oligonucleotides for
XX antisense inhibition of gene expression useful e.g. for treating tumors
XX but the oligonucleotides may also be directed against genes present
XX in pathogens (e.g. Plasmodium or viruses/viroids, pathogenic on humans,
XX animals or plants) or against cytokine, Id, developmental or prion genes.
XX The method provides more effective inhibition of gene expression than use
XX of known oligonucleotides, probably because the unpaired overhang
XX increases stability and thus intracellular concentration
XX
XX Sequence 3417 BP; 645 A; 1047 C; 1062 G; 663 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16; DB 5; Length 3417;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 GCGCCAGCCGCCACAA 20
Db 657 GCGCCAGCCGCCACAA 642
RESULT 22
AAQ47064/c
ID AAQ47064 standard; cDNA; 3845 BP.
XX
XX AC AAQ47064;
XX
XX 25-MAR-2003 (revised)
DT 12-JAN-1994 (first entry)
XX
XX tie receptor kinase cDNA.
XX
XX Tie; receptor; tyrosine kinase; endothelial cell; immunoglobulin; Ig;
XX epidermal growth factor; EGF; repeat; blood vessel; atherosclerosis;
XX thromboembolic disease; neoplastic disease; tumour angiogenesis;
XX wound healing; inflammatory disease; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 37..3453
XX sig_peptide /tag= a
XX mat_peptide /tag= b
XX /tag= c
XX
XX WO9314124-A1.
XX
XX 22-JUL-1993.
XX
XX 08-JAN-1993; 93WO-FI000006.
XX
XX 09-JAN-1992; 92US-00817800.
XX
XX (UYHE-) UNIV HELSINKI HOLDING LTD.
XX
XX Partanen J, Armstrong E, Makela TP, Korhonen J, Alitalo K;
XX
```

```
DR WPI; 1993-243152/30.
DR P-ESDB; AAR39820.
XX
XX Nucleic acid encoding tyrosine kinase receptor - for regulating tie
XX activity in diagnosis and treatment of neoplastic diseases involving
XX tumour angiogenesis, wound healing etc.
XX
XX Claim 2; Page 50-53; 73pp; English.
XX
XX The sequence given in AAQ47064 represents the full length 'tie' receptor
XX kinase cDNA. The sequence given in AAQ47065 represents a truncated
XX version of this, wherein nucleotides corresponding to positions 676- 807
XX of the first sequence are absent. Tie is an endothelial cell receptor
XX tyrosine kinase and contains immunoglobulin- (Ig-) and epidermal growth
XX factor- (EGF-) like repeats. The tie gene is expressed in the endothelial
XX cells of blood vessels. Tie levels may indicate onset/progression of
XX diseases involving endothelial cells and their tie receptors, such as
XX neoplastic diseases involving tumour angiogenesis, wound healing,
XX thromboembolic diseases, atherosclerosis and inflammatory diseases.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16; DB 2; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 GCGCCAGCCGCCACAA 20
Db 693 GCGCCAGCCGCCACAA 678
RESULT 23
ABZ34892/c
ID ABZ34892 standard; cDNA; 3845 BP.
XX
XX AC ABZ34892;
XX
XX 05-FEB-2003 (first entry)
DT
XX
XX Human gene expression profile polynucleotide SEQ ID NO 4.
XX
XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
XX bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
XX tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
XX gene expression; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200274979-A2.
XX
XX 26-SEP-2002.
XX
XX 20-MAR-2002; 2002WO-US008456.
XX
XX 20-MAR-2001; 2001US-0276947P.
XX
XX (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.
XX
XX Wan J, Wang Y;
XX
XX WPI; 2002-740862/80.
XX
XX New gene expression profile generated from primary, endothelial,
XX epithelial, and muscle cell types, useful for identifying disease
XX pathologies involving alterations of gene expression, e.g. cancer.
XX
XX Claim 1; Page 223-224; 850pp; English.
XX
XX The invention relates to a gene expression profile comprising one or more
XX genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
XX is a coronary artery endothelium, umbilical artery or vein endothelium,
XX aortic endothelium, dermal microvascular endothelium, pulmonary artery
```


endothelium, myometrium microvascular endothelium, keratinocyte
epithelium, bronchial epithelium, mammary epithelium, prostate
epithelium, renal cortical epithelium, renal proximal tubule epithelium,
small airway epithelium, renal epithelium, umbilical artery smooth
muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
osteoblasts or prostate stromal cell. The gene expression profile is used
for determining the level of RNA expression for a sample, determining the
phenotype of a cell and distinguishing cell types. The gene or a protein
expression profile is useful in identifying disease pathologies involving
alterations of gene expression. The assessment of expression profiles may
provide meaningful information with respect to tumour type and stage,
treatment methods, and prognosis. The gene or protein expression profile
may also be used for creating microarrays. The microarray is useful for
genetic and physical mapping of genomes, DNA sequencing, genetic or
medical diagnosis, genotyping of organisms, confirming cell or tissue
identifications and in identifying promising antibiotics, antiviral or
antifungal agents

XX Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 6; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCGACAA 20
Db 693 GCGCCAGCCGACAA 678

RESULT 24
ABV94119/c
ID ABV94119 standard; cDNA; 3845 BP.
XX
AC ABV94119;
XX
DT 08-JAN-2003 (first entry)
XX
DE Breast carcinoma related nucleotide sequence SEQ ID NO:110.
XX
KW Human; breast carcinoma; cancer; tumour; cytostatic; anti-tumour; gene;
KW ss.
XX Homo sapiens.
XX WO200246467-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-1B002811.
XX
PR 08-DEC-2000; 2000US-0254090P.
PR 07-DEC-2001; 2001US-00007926.
XX
PA (IPSO-) IPSOGEN.
XX
PI Bertucci F, Houlgatte R, Birnbaum D, Nguyen C, Viena P, Fert V;
XX WPI; 2002-619023/66.
XX

Novel polynucleotide library useful in molecular characterization of a
carcinoma, comprising a pool of polynucleotide sequences or its
subsequences which are either underexpressed or overexpressed in tumor
cells.

Claim 1; Page 181-182; 401pp; English.

The present invention describes a polynucleotide library (I) useful in
the molecular characterization of a carcinoma, comprising a pool of
polynucleotides or its subsequences which are either underexpressed or
overexpressed in tumour cells, and correspond to any of the

polynucleotide sequences chosen from the 468 sequences given in ABV94010
to ABV94477. Also described: (1) a polynucleotide array (II) useful for
the prognosis or diagnostic of tumour, comprising (I); and (2) detecting
(M1) differentially expressed polynucleotide sequences which are
correlated with a cancer, involves obtaining a polynucleotide sample from
a patient, and reacting the polynucleotide sample obtained with a probe
immobilised on a solid support, where the probe comprises any combination
of the polynucleotide sequences of (I) or its expression products encoded
by polynucleotide sequences of (I), and detecting the reaction product.
CC (I) have cytostatic activities and can be used as anti-tumour agents. (I)
is useful in molecular characterisation of a carcinoma. (I) and (II) are
useful for the prognosis or diagnostic of tumour, in differentiating a
normal cell from a cancer cell, detecting a hormone sensitive tumour
cell, differentiating a tumour with lymph nodes from a tumour without
lymph nodes, differentiating antracycline-sensitive tumours from
antracycline-insensitive tumours, and classifying good and poor prognosis
primary breast tumors. (I) is useful for large-scale molecular
characterisation of breast cancer that help in prediction, prognosis and
cancer treatment, and for detecting differentially expressed genes that
correlated with a cancer

XX Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 6; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCGACAA 20
Db 693 GCGCCAGCCGACAA 678

RESULT 25
ABX08795/c
ID ABX08795 standard; cDNA; 3845 BP.
XX
AC ABX08795;
XX
DT 21-JAN-2003 (first entry)
XX
DE Angiogenesis-associated human polynucleotide sequence #57.
XX
KW Human; angiogenesis-associated transcript; angiogenesis;
KW angiogenesis-associated disease; cancer; cytostatic; gene therapy; gene;
KW ss.
XX Homo sapiens.
XX WO200279492-A2.
XX
PD 10-OCT-2002.
XX
PF 14-FEB-2002; 2002WO-US0004915.
XX
PR 14-FEB-2001; 2001US-00784356.
PR 22-FEB-2001; 2001US-00791390.
PR 19-APR-2001; 2001US-0285475P.
PR 03-AUG-2001; 2001US-0310025P.
PR 13-NOV-2001; 2001US-0350666P.
PR 29-NOV-2001; 2001US-0334244P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.

Murray R, Glynn R, Watson SR, Aziz N;

WPI; 2003-040681/03.

P-PSDB; ABU03511.

Detecting angiogenesis-associated transcript in a cell for diagnosing and
treating cancer by contacting a sample with a polynucleotide that
exhibits changes in expression level as a function of time in tissue
undergoing angiogenesis.

PS Example 2; Page 232-233; 291pp; English.

XX The present invention relates to methods and compositions for detecting
CC an angiogenesis-associated transcript in a cell in a patient. The method
CC involves contacting a biological sample from the patient with a
CC polynucleotide that selectively hybridises to a sequence at least 80%
CC identical to any of the angiogenesis-associated human polynucleotide
CC sequences given in the specification. These angiogenesis-associated
CC polynucleotide sequences comprise genes that exhibit changes in
CC expression levels as a function of time in tissue undergoing
CC angiogenesis. The method and the polynucleotide sequences of the
CC invention are useful for diagnosing and treating angiogenesis and
CC angiogenesis-associated diseases e.g. cancer. The polynucleotide
CC sequences are also useful in the gene therapy of such disorders. The
CC angiogenesis-associated proteins encoded by the polynucleotide sequences
CC are useful as a vaccine for therapeutic and prophylactic immunisation.
CC ABX08739-ABX08853 represent angiogenesis-associated polynucleotide
XX sequences

SQ Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 7; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GCGCCCGAGCCCAAA 20
| | | | | | | | | |
DB 693 GCGCCCGAGCCCAAA 678

RESULT 26

ABX12538/c
ID ABX12538 standard; cDNA; 3845 BP.

AC ABX12538;

DT 10-MAY-2003 (first entry)

DE cDNA encoding human Tie receptor tyrosine kinase 1 (Tie 1).

XX Cytostatic; vasodilator; antiinflammatory; cardiant; gene therapy;
KW ligand-receptor binding modulator; ephrin ligand; angiogenesis;
KW lymphangiogenesis; aberrant Ephrin-Tie biology; cell growth disorder;
KW cell migration disorder; cell proliferation disorder; neovascularisation;
KW ischaemia; infarction; tissue graft; transplant; human; tie 1;
KW tie receptor tyrosine kinase 1; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 37..3453
FT /tag= a
FT /product= "Tie 1"
FT /note= "Tie receptor tyrosine kinase"

PN WO2003004529-A2.

PD 16-JAN-2003.

PF 02-JUL-2002; 2002WO-IB002524.

PR 02-JUL-2001; 2001US-0302960P.

PA (LICN) LICENTIA LTD.

PI Alitalo K, Kubo H;

XX WPI; 2003-210341/20.

DR P-PSDB; ABU07837.

XX Identifying modulators of binding between a Tie receptor tyrosine kinase
PT and an Ephrin ligand, useful for promoting neovascularization, comprises
PT contacting a Tie receptor with an Ephrin in the presence of a putative

PT modulator.

XX Disclosure; Page 56-61; 199pp; English.

XX The invention describes a method of identifying a modulator of binding
CC between a Tie receptor tyrosine kinase and an Ephrin ligand. The method
CC comprises contacting a Tie receptor composition with an Ephrin
CC composition in the presence and in the absence of a putative modulator
CC compound, and detecting the binding between the Tie receptor and the Ephrin
CC in the presence and in the absence of the putative modulator. The method
CC is useful for identifying a modulator of binding between a Tie receptor
CC tyrosine kinase and an Ephrin ligand. Modulators identified from the
CC method are useful in modulating angiogenic processes, including
CC lymphangiogenesis, for treating diseases associated with aberrant Ephrin-
CC Tie biology, aberrant growth, migration or proliferation of cells that
CC express a Tie receptor, or for promoting growth of vessel or
CC neovascularisation (e.g. ischaemic tissue, an infarction, a new or
CC chronic compound, or a tissue graft or transplant). This sequence encodes
XX human Tie receptor tyrosine kinase 1 (Tie 1)

SQ Sequence 3845 BP; 743 A; 1179 C; 1155 G; 768 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 7; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GCGCCCGAGCCCAAA 20
| | | | | | | | | |
DB 693 GCGCCCGAGCCCAAA 678

RESULT 27

ABX63148/c
ID ABX63148 standard; cDNA; 3914 BP.

AC ABX63148;

DT 25-FEB-2003 (first entry)

DE Human cDNA #148 differentially expressed in activated vascular tissue.

XX Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW gene therapy; vascular disease; cancer; coronary; artery disease;
KW hypertension; diabetes; pre-eclampsia; restenosis;
KW ischaemia-reperfusion injury; stroke.

XX Homo sapiens.

XX US2002137081-A1.

XX 26-SEP-2002.

XX 08-JAN-2002; 2002US-00044090.

XX 28-JUL-2000; 2000US-0222469P.

XX 08-JAN-2001; 2001US-0260483P.

XX (BAND/) BANDMAN O.

XX Bandman O;

XX WPI; 2003-110597/10.

XX Combination for diagnosing, staging, treating, or monitoring the
PT progression of treatment of a vascular disease, e.g. atherosclerosis,
PT comprises several cDNAs that are differentially expressed in activated
PT vascular tissue.

XX Claim 1; Page; 18pp; English.

XX This invention relates to a combination comprising several cDNAs that are
CC differentially expressed in activated vascular tissue. The invention also

CC discloses a high throughput method for detecting differentially expressed
 CC cDNAs in a sample. The cDNAs of the invention may have
 CC antiarteriosclerotic, cytotatic, cardiac, hypotensive, antidiabetic,
 CC gynaecological, vasotropic and cerebroprotective activities and may be
 CC used in gene therapy. The cDNAs of the invention may be used in a high-
 CC throughput methods for detecting differential expression of one or more
 CC cDNAs in a sample, or screening several molecules or compounds to
 CC identify a molecule or compound that specifically binds a cDNA of the
 CC invention. A protein encoded by the cDNA may be used to screen several
 CC molecules or compounds to identify a ligand that specifically binds to
 CC the protein, or to produce or purify an antibody to the protein that can
 CC be used to detect a protein in a sample or purify a natural or
 CC recombinant protein from a sample. The nucleotides may be useful for
 CC diagnosing, staging, treating, or monitoring the progression of treatment
 CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
 CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
 CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
 CC genetic or gene expression analysis of several new nucleic acid
 CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
 CC diagnosing pre-pathologic disorders, and chronic or acute diseases
 CC associated with abnormalities in the expression, amount or distribution
 CC of the protein. The present sequence represents a cDNA of the invention
 CC that is differentially expressed in activated vascular tissue. Note: the
 CC sequence data for this patent did not form part of the specification, but
 CC was obtained in electronic format directly from USPTO at
 CC <http://seqdata.uspto.gov/sequence.html?docID=20020137081>
 XX
 XX Sequence 3914 BP; 768 A; 1205 C; 1168 G; 773 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 7; Length 3914;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCCGAGCCCCACAA 20
 DB 754 GCGCCCGAGCCCCACAA 739

RESULT 28

ACH03957/c

ID ACH03957 standard; cDNA; 3914 BP.

XX ACH03957;

XX 26-SEP-2003 (first entry)

XX Human cDNA differentially expressed in lung cancer #162.

XX Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
 XX .respiratory disorder; lung cancer; asthma; human.

XX Homo sapiens.

XX US2003065157-A1.

XX 03-APR-2003.

XX 04-APR-2002; 2002US-00116802.

XX 04-APR-2001; 2001US-0281593P.

XX (LASE/) LASEK A W.

XX Lasek AW;

XX WPI; 2003-540803/51.

XX New combination comprising cDNAs that are differentially expressed in
 XX respiratory disorders, useful for diagnosing or treating respiratory
 XX disorders e.g., lung cancer, chronic obstructive pulmonary disease,
 XX emphysema or asthma.

XX Claim 1; Page; 39pp; English.

XX The invention relates to a combination comprising cDNAs or their
 CC complements that are differentially expressed in respiratory disorder.
 CC The combination is useful for preparing a composition for diagnosing or
 CC treating respiratory disorders e.g. lung cancer, chronic obstructive
 CC pulmonary disease, emphysema or asthma. The present sequence represents
 CC human cDNA differentially expressed during lung cancer

XX Sequence 3914 BP; 768 A; 1205 C; 1168 G; 773 T; 0 U; 0 Other;

Query Match 80.0%; Score 16; DB 8; Length 3914;

Best Local Similarity 100.0%; Pred. No. 1.2e+03;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCCGAGCCCCACAA 20

DB 754 GCGCCCGAGCCCCACAA 739

RESULT 29

AAZ32059/c

ID AAZ32059 standard; DNA; 8670 BP.

XX AAZ32059;

XX 10-JAN-2000 (first entry)

XX Human METH2 related EST X16619.

XX Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
 XX cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 XX angiogenesis inhibitor; abnormal wound healing; inflammation;
 XX rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 XX diabetic retinopathy; macula degeneration; haemangioma; detection;
 XX arterial-venous malformation; immune deficiency; ss.

XX Homo sapiens.

XX WO9937660-A1.

XX 29-JUL-1999.

XX 22-JAN-1999; 99WO-US001313.

XX 23-JAN-1998; 98US-0072298P.

XX 28-AUG-1998; 98US-0098539P.

XX (IRUE/) IRUELA-ARISPE L.

XX (HAST/) HASTINGS G A.

XX (RUBE/) RUBEN S M.

XX Iruela-Arispe L, Hastings GA, Ruben SM;

XX WPI; 1999-590684/50.

XX New isolated metalloprotease thrombospondin polypeptides, useful for
 XX treating hyperproliferative disorders, cancers or autoimmune disorders.

XX Disclosure; Page 431-437; 457pp; English.

XX AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
 CC respectively. METH1 and METH2 have been found to be potent inhibitors of
 CC angiogenesis both in vitro and in vivo. They can be used for treating
 CC cancer and other disorders related to angiogenesis including abnormal
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis, endometrial
 CC bleeding disorders, diabetic retinopathy, some forms of macula
 CC degeneration, haemangiomas, and arterial-venous malformations. They may
 CC be useful in treating deficiencies or disorders of the immune system, by
 CC activating or inhibiting the proliferation, differentiation, or
 CC mobilisation (chemotaxis) of immune cells. The etiology of these immune
 CC deficiencies or disorders may be genetic, somatic, such as cancer or some
 CC autoimmune disorders, acquired (e.g. by chemotherapy or toxins), or

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 934.603 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-11

Perfect score: 20

Sequence: 1 cgaggcgccagcccccacaa 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pln:**

20: em_gss_prt:**

21: em_gss_fun:**

22: em_gss_mam:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 18.4 | 92.0 | 541 | 9 | AU192152 |
| C 2 | 17.4 | 87.0 | 144 | 9 | AZ207487 |
| C 3 | 17.4 | 87.0 | 183 | 12 | BI298853 |
| C 4 | 17.4 | 87.0 | 221 | 10 | BF333695 |
| | | | | | ----- |
| | | | | | AU192152 AU192152 |
| | | | | | AZ207487 mv89g06.r |
| | | | | | BI298853 UI-R-CV2- |
| | | | | | BF333695 QV0-CS000 |


```
224 15.8 79.0 430 9 AI802084
225 15.8 79.0 431 12 BM840118
226 15.8 79.0 432 14 A2039431
227 15.8 79.0 433 14 H65577
228 15.8 79.0 434 10 AW414902
229 15.8 79.0 435 10 BE674158
230 15.8 79.0 436 14 W8068
231 15.8 79.0 437 10 BE468248
232 15.8 79.0 438 14 W39608
233 15.8 79.0 439 10 FR0038212
234 15.8 79.0 440 14 CF555178
235 15.8 79.0 441 28 BZ608882
236 15.8 79.0 442 14 CB785676
237 15.8 79.0 443 9 AI087321
238 15.8 79.0 444 14 H70038
239 15.8 79.0 445 9 AI364711
240 15.8 79.0 446 12 BI748915
241 15.8 79.0 447 9 AI310009
242 15.8 79.0 448 29 CE762574
243 15.8 79.0 449 9 AI236897
244 15.8 79.0 450 28 AZ589021
245 15.8 79.0 451 13 BX519576
246 15.8 79.0 452 13 BQ143077
247 15.8 79.0 453 14 CB142053
248 15.8 79.0 454 28 AZ754830
249 15.8 79.0 455 10 BE09873
250 15.8 79.0 456 28 AQ339508
251 15.8 79.0 457 10 BB759368
252 15.8 79.0 458 14 CD545362
253 15.8 79.0 459 9 AI160577
254 15.8 79.0 460 12 BM687393
255 15.8 79.0 461 14 CD538335
256 15.8 79.0 462 9 AI433386
257 15.8 79.0 463 10 AW553832
258 15.8 79.0 464 9 AI225675
259 15.8 79.0 465 10 AW671976
260 15.8 79.0 466 9 AI154624
261 15.8 79.0 467 9 AI962454
262 15.8 79.0 468 10 BE973671
263 15.8 79.0 469 12 BQ017350
264 15.8 79.0 470 28 AZ200586
265 15.8 79.0 471 9 AI047721
266 15.8 79.0 472 13 BX511881
267 15.8 79.0 473 10 BE878880
268 15.8 79.0 474 9 AI306825
269 15.8 79.0 475 12 BM701338
270 15.8 79.0 476 10 AW674953
271 15.8 79.0 477 10 BE232321
272 15.8 79.0 478 10 AW953545
273 15.8 79.0 479 10 BF133796
274 15.8 79.0 480 13 BQ189194
275 15.8 79.0 481 10 AW139551
276 15.8 79.0 482 14 CD542765
277 15.8 79.0 483 13 BQ830017
278 15.8 79.0 484 14 CB158908
279 15.8 79.0 485 14 CD542751
280 15.8 79.0 486 12 BQ061438
281 15.8 79.0 487 10 CF534607
282 15.8 79.0 488 29 CE156722
283 15.8 79.0 489 9 AI131161
284 15.8 79.0 490 9 AI154675
285 15.8 79.0 491 14 CD538158
286 15.8 79.0 492 10 BE312164
287 15.8 79.0 493 29 CE069728
288 15.8 79.0 494 13 BX668561
289 15.8 79.0 495 12 BM352370
290 15.8 79.0 496 14 CB114642
291 15.8 79.0 497 9 AI673892
292 15.8 79.0 498 10 BE871180
293 15.8 79.0 499 10 BE621823
294 15.8 79.0 500 10 AW272400
295 15.8 79.0 501 9 AI043584
296 15.8 79.0 502 10 AW248200
```

```
AI802084 tx24f02.x
BM840118 K-EST0117
A2039431 RPI-23-3
H65577 yr52a09.r1
AW414902 48905.MAR
BE674158 7d76a05.x
W8068 2d72e07.r1
BE468248 hz71b02.x
W39608 zc19e04.r1
AI125713 Fugu.rubr
CF555178 1115006A1
BZ608882 WHA8497F
CB785676 AMGN09C1M
AI087321 oz45f09.x
H70038 yu73b11.r1
AI364711 qz07b05.x
BI748915 rc083f11.y
AI310009 EST180922
CE762574 tigr-gss-
AA236897 z94309.s
AZ589021 IM0397P07
BX519576 BX519576
BQ143077 fmb1c.pk0
CB142053 K-EST0195
AZ754830 CQ04802.f
BE09873 RCS-BT070
AQ339508 HS 5022.A
BB759368 BB759368
CD545362 BQ266C01-
AI160577 qc83c04.x
BM687393 UI-E-CQ1-
CD538335 B020C11-
AI433386 t973a06.x
AW553832 L0232D01-
AI225675 uj14b12.y
AW671976 LGI 353.D
AI154624 AU154624-
AI962454 wq52e06.x
BE973671 601690849
BQ017350 UI-H-D11-
AZ200586 SP 1039.A
AI047721 DKE2p860
BX511881 BX511881
BE878880 601493175
AI306825 EST177820
BM701338 UI-E-EJ0-
AW674953 B860C02.y
BE232321 137313.WA
AW953545 EST365615
BF133796 601900471
BQ189194 UI-E-EJ1-
AW139551 UI-H-B11-
CD542765 B0243E07-
BQ830017 Lf61n2041
CB158908 K-EST0218
CD542751 B0243D10-
BQ061438 BQ061438
CF534607 UI-M-G10-
CE156722 tigr-gss-
AI131161 qc15c01.x
AI154675 AU154675
CD538158 B0200G10-
BE312164 601152459
CE069728 tigr-gss-
BX668561 BX668561
BM352370 i989007.Y
CB114642 K-EST0154
AL673892 AL673892
BE871180 601448752
BE621823 601494075
AW272400 xv14b03.x
AI043584 DKFZ434F
AW248200 2819721.5
```

297 15.8 79.0 592 12 BI796574
298 15.8 79.0 593 14 CB269034
299 15.8 79.0 593 29 CE442639
300 15.8 79.0 596 10 AW553048

ALIGNMENTS

RESULT 1
AA192152/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 541
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PFL026f01.z"
/dev_stage="sporophytes"
/clone_lib="Porphyra yezoensis TU-1 sporophytes"

ORIGIN
Query Match 92.0%; Score 18.4; DB 9; Length 541;
Best Local Similarity 95.0%; Pred. No. 3.5e+03;
Matches: 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CGAGCGCCCGAGCCCAACAA 20
|||||
Db 156 CGAGCGCCCGAGCCCAACAA 137

RESULT 2
AA207487/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AA207487
mv89906.r1 GuayWoodford Beier mouse kidney day 7 Mus musculus cdna
clone IMAGE:662266 5', mRNA sequence.
AA207487
AA207487.1 GI:1804059
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 144)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project
WASHU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:408114

Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 133.

FEATURES

source

```
1. .144
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:662266"
/tissue_type="kidney"
/dev_stage="juvenile (7 days old)"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="GuayWoodford Beier mouse kidney day 7"
/notes="Organ: kidney; Vector: pBluescript SK-; Site 1:
ECORI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTCCTTTTTCCTTTT 3' Library provided
by Lisa Guay-Woodford."
```

ORIGIN

```
Query Match      87.0%; Score 17.4; DB 9; Length 144;
Best Local Similarity 94.7%; Pred. No. 6.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CGAGCGCCCGCCGCCCCACA 19
Db 108 CGAGCGCCCGCCGCCCCACA 90
```

RESULT 3

```
BI298853/c
LOCUS      183 bp mRNA linear EST 20-JUL-2001
DEFINITION UI-R-CV2-chv-d-06-0-UI-s1 UI-R-CV2 Rattus norvegicus cDNA clone
ACCESSION BI298853
VERSION BI298853.1 GI:14975133
KEYWORDS EST.
```

SOURCE

Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 183)
Bonaldo,M.P., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to verify it as a clone from the
normalized rat eye library cDNA library Preparation: M.B. Soares

Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

```
1. .183
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CV2-chv-d-06-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CV2"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; CV2 is a
subtracted library derived from a normalized eye library
(mRE2), constructed according to the procedure described
by Bonaldo, Lennon & Soares (Genome Research Genome 6:
791-806, 1996). For construction of the CV2 library,
plasmid DNA from the normalized rat eye library was
electroporated into competent bacteria for production of
single-stranded circular DNA. This was then used as a
tracer in a subtractive hybridization with a driver
comprising: a) a set of about 1,000 arrayed clones from
the non-normalized eye library CV0 (PCR amplified inserts
from a plasmid DNA template preparation representing
plates R-CV0-BRH through R-CV0-BRR). This represented 20%
of the final driver population. b) A set of about 4,000
arrayed clones from the normalized eye library CV1 (PCR
amplified inserts from a plasmid DNA template preparation
representing plates R-CV1-BRS through R-CV1-BSC, R-CV1-BSE
through R-CV1-BTC, and R-CV1-BV0 through R-CV1-BVU). This
represented 80% of the final driver population.
TAG TISSUE=rat eye
TAG_LIB=UI-R-CV2
TAG_SEQ=CAGCC"
```

ORIGIN

```
Query Match      87.0%; Score 17.4; DB 12; Length 183;
Best Local Similarity 94.7%; Pred. No. 6.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 GAGCGCCCGCCGCCCCACAA 20
Db 82 GGGCGCGCCGCCCCACAA 64
```

RESULT 4

```
BF333695
LOCUS      221 bp mRNA linear EST 22-NOV-2000
DEFINITION QV0-CS0009-220900-401-h07 CS0009 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF333695
VERSION BF333695.1 GI:11304443
KEYWORDS EST.
```

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 221)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zagor,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.U.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-CS00009-
220900-401-h07&t3=2000-09-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 221.

FEATURES
source

1. .221
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CS0009"
note="Organ: colon_est; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 221;
Best Local Similarity 94.7%; Pred. No. 7.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGGCCCCAGCCCAAA 20
|||||
DB 107 GAGGGCCCCAGCCCAAA 125

RESULT 5
AI713710/c

LOCUS AI713710 222 bp mRNA linear EST 08-JUN-1999
DEFINITION UI-R-AGI-aao-b-11-0-UI.s1 UI-R-AGI Rattus norvegicus cDNA clone
UI-R-AGI-aao-b-11-0-UI 3', mRNA sequence.
ACCESSION AI713710
VERSION AI713710.1 GI:5017510
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 222)
Ronaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
8889548
PUBMED
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized ventricle at 13 dpc library cDNA Library Preparation:
M.B. Soares Lab clone distribution: clones will be available
through Research Genetics (www.resgen.com)

Seq primer: M13 Forward
POLYA=yes.

FEATURES
source

1. .222
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-AGI-aao-b-11-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-AGI"
note="Vector: pTVT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-AGI
library is a normalized library constructed from 13 dpc
rat ventricle. The tag is a string of 6 nucleotides
present between the Not I site and the oligo-dT track.
The library was constructed as described by Bonaldo,
Lennon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Jim Lin, Department of Biology,
University of Iowa.
TAG_TISSUE=ventricle at 13 dpc
TAG_LIB=UI-R-AGI
TAG_SEQ=CAGCGA"

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 222;
Best Local Similarity 94.7%; Pred. No. 7.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGGCCCCAGCCCAAA 20
|||||
DB 73 GGGGGCCCCAGCCCAAA 55

RESULT 6
AA657083/c

LOCUS AA657083 245 bp mRNA linear EST 04-NOV-1997
DEFINITION vs23c10.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
IMAGS:1139058 5', mRNA sequence.

ACCESSION
VERSION

AA657083
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 245)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

TITLE
JOURNAL

COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MGI:620330
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 224.

FEATURES
source

1. .245
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H"
/db_xref="taxon:10090"

```

/clone="IMAGE:1139058"
/cell_line="C2C12"
/lab_host="DH108"
/clone_lib="Barstead mouse myotubes MPLRB5"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoRI, Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCAGATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCCTTG], digested with Not I and cloned into the
Library constructed by Bob Barstead. The C2C12 cell line
(avaiable from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins. "
ORIGIN
Query Match 87.0%; Score 17.4; DB 9; Length 245;
Best Local Similarity 94.7%; Pred. No. 7.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCA 19
||||| ||||| |||||
Db 187 CGAGGCGCCAGCCCA 169

RESULT 7
AA637917/c
LOCUS
DEFINITION
v33h02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
IMAGE:1122483 5', mRNA sequence.

ACCESSION
AA637917
VERSION
AA637917.1 GI:2561505
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 281)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HHMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1611819
Putative full length read
vector to vector length is 615
Seq primer: -28m13 rev2 ET from AmerSham
High quality sequence stop: 185.
Location/Qualifiers
1..281
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:1122483"
/cell_line="C2C12"
/lab_host="DH108"
/clone_lib="Barstead mouse myotubes MPLRB5"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCAGATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTT
3']; on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified

```

```

TGTTCAGATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(avaiable from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins. "
ORIGIN
Query Match 87.0%; Score 17.4; DB 9; Length 281;
Best Local Similarity 94.7%; Pred. No. 7.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCA 19
||||| ||||| |||||
Db 156 CGAGGCGCCAGCCCA 138

RESULT 8
W83781/c
LOCUS
DEFINITION
mf32h08.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
Clone IMAGE:406815 5', mRNA sequence.

ACCESSION
W83781
VERSION
W83781.1 GI:1541010
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 283)
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HHMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
On Sep 12, 1996 this sequence version replaced gi:1394734.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:250583
Seq primer: -28m13 rev2 from AmerSham
High quality sequence stop: 268.
Location/Qualifiers
1..283
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:406815"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH108"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCAGATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTTTTT
3']; on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified

```

PTT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 283;
Best Local Similarity 94.7%; Pred. No. 7.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19
|||||
Db 167 CGAGGCGCCAGCCCCACA 149
|||||

RESULT 9
AA623516/c
LOCUS
DEFINITION
v33ho5.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
clone IMAGE:1023033 5', mRNA sequence.

ACCESSION
AA623516
VERSION
AA623516.1 GI:2527392
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

REFERENCE
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE
JOURNAL
COMMENT
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:573809
High quality sequence stop: 269.

FEATURES
source
1..304
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J x DBA/2J F1"
/db_xref="taxon:10090"
/clone="IMAGE:1023033"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/clone_lib="Knowles Solter mouse blastocyst B1"
/note="Organ: embryo; Vector: pSPORT; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI (dT):
5'-CGGTCGACCGCGACCGTTTTTTTTTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 304;
Best Local Similarity 94.7%; Pred. No. 7.7e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19
|||||
Db 91 CGAGGCGCCAGCCCCACA 73
|||||

RESULT 10
AA144304/c
LOCUS
DEFINITION
mr97f10.r1 Stratagene mouse embryonic carcinoma (#937317) Mus
musculus cDNA clone IMAGE:605419 5', mRNA sequence.

ACCESSION
AA144304
VERSION
AA144304.1 GI:1713674
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

REFERENCE
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE
JOURNAL
COMMENT
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:370851
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 262.

FEATURES
source
1..305
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:605419"
/tissue_type="carcinoma"
/dev_stage="embryonic"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse embryonic carcinoma
(#937317)"
/note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. P19 cell
line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 305;
Best Local Similarity 94.7%; Pred. No. 7.7e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCAGCCCCACA 19
|||||
Db 181 CGAGGCGCCAGCCCCACA 163
|||||

RESULT 11
BQ301646/c
LOCUS
DEFINITION
MRI-HB0017-151200-003-d07 HB0017 Homo sapiens cDNA, mRNA sequence.

ACCESSION
BQ301646
VERSION
BQ301646.1 GI:20817168
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 308)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663
MEDLINE
10737800
PUBMED

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR1&t2=MR1-HB0017-151200-003-d07&t3=2000-12-15&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 6.

Location/Qualifiers

FEATURES

source
1..308
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HB0017"

/note="Organ: bocio tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 308;
Best Local Similarity 94.7%; Pred. No. 7.7e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCACAA 20
|||||
Db 208 GAGGCGCCCGCCACAA 190

RESULT 12
BQ301647/c 312 bp mRNA linear EST 16-MAY-2002
LOCUS
DEFINITION
MR1-HB0017-151200-003-e03 HB0017 Homo sapiens cDNA, mRNA sequence.

ACCESSION BQ301647

VERSION BQ301647.1 GI:20817169

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 312)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663
MEDLINE
10737800
PUBMED

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR1&t2=MR1-HB0017-151200-003-e03&t3=2000-12-15&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 38.

FEATURES

Location/Qualifiers

1..312

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HB0017"

/note="Organ: bocio tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 87.0%; Score 17.4; DB 13; Length 312;
Best Local Similarity 94.7%; Pred. No. 7.7e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCACAA 20
|||||
Db 212 GAGGCGCCCGCCACAA 194

RESULT 13

AI137544/c

LOCUS

DEFINITION

UI-R-C2p-oc-g-03-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone

UI-R-C2p-oc-g-03-0-UI 3', mRNA sequence.

ACCESSION AI137544

VERSION AI137544.1 GI:3638321

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 316)

Ronaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Ovary library. cDNA Library Preparation: M. Faima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Consortium at LNL (info@image.lnl.gov). IMAGE ID=1787271
Seq primer: M13 Forward
POLYA=No.

[illegible]


```

ORIGIN
  Query Match      87.0%; Score 17.4; DB 9; Length 370;
  Best Local Similarity 94.7%; Pred. No. 8e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCCCCACA 19
    ||||| ||||| ||||| ||||| |||||
Db 109 CGAGGCACCCAGCCCCCACA 91

RESULT 18
A1180711/c
LOCUS ub1808.r1 Soares mammary gland MbMMG Mus musculus cDNA clone
DEFINITION IMAGE:1395831 5', mRNA sequence.
ACCESSION A1180711
VERSION A1180711.1 GI:3731349
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:907547
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 307.

FEATURES
Location/Qualifiers
  1..396
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="CS7BL/6J"
    /db_xref="taxon:10090"
    /clone="IMAGE:1395831"
    /sex="male"
    /tissue_type="mammary gland"
    /dev_stage="4 weeks"
    /lab_host="DH10B"
    /clone_lib="Soares mammary gland MbMMG"
    /note="Organ: mammary gland; Vector: pT73D-Pac
    (Pharmacia) with a modified polylinker; Site 1: Not I;
    Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
    oligo(dT) primer [5'
    TGTTACCAATCTGAAGTGGAGCGCGCGAGTGTGTTTTTTTTTTTTTTTTTTT
    T 3']; double-stranded cDNA was ligated to Eco RI
    adaptors (Pharmacia), digested with Not I and cloned into
    the Not I and Eco RI sites of the modified pT73 vector.
    RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
    constructed and normalized by Bento Soares and M.Patima
    Bonaudo."

ORIGIN
  Query Match      87.0%; Score 17.4; DB 9; Length 396;
  Best Local Similarity 94.7%; Pred. No. 8.1e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCCCCACA 19
    ||||| ||||| ||||| ||||| |||||
Db 109 CGAGGCACCCAGCCCCCACA 91

RESULT 18
A1180711/c
LOCUS ub1808.r1 Soares mammary gland MbMMG Mus musculus cDNA clone
DEFINITION IMAGE:1395831 5', mRNA sequence.
ACCESSION A1180711
VERSION A1180711.1 GI:3731349
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:907547
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 307.

FEATURES
Location/Qualifiers
  1..432
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /clone="IMAGE:1617966"
    /tissue_type="hypothalamus"
    /lab_host="DH10B"
    /clone_lib="Soares mouse hypothalamus NMHy"
    /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
    modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
    strand cDNA was primed with a Not I - oligo(dT) primer [5'
    TGTTACCAATCTGAAGTGGAGCGCGCGAGTGTGTTTTTTTTTTTTTTTTTTT
    T 3']; double-stranded cDNA was ligated to Eco RI adaptors
    (Pharmacia), digested with Not I and cloned into the Not I
    and Eco RI sites of the modified pT73 vector. RNA
    provided by Dr. Wolfgang Liedtke. Library went through
    two rounds of normalization, and was constructed by Bento
    Soares and M.Patima Bonaudo."

ORIGIN
  Query Match      87.0%; Score 17.4; DB 9; Length 432;
  Best Local Similarity 94.7%; Pred. No. 8.3e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCCCCACA 19
    ||||| ||||| ||||| ||||| |||||
Db 193 CGAGGCACCCAGCCCCCACA 175

RESULT 20
BG373356/c
LOCUS BG373356
DEFINITION UI-R-CV1-brx-g-10-0-UI.s1 UI-R-CV1 Rattus norvegicus cDNA clone
ACCESSION BG373356
VERSION BG373356.1 GI:13269893
  
```

KEYWORDS SOURCE ORGANISM

EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE AUTHORS TITLE

1 (bases 1 to 434)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL MEDLINE PUBMED COMMENT

97044477
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to verify it as a clone from the
normalized rat eye library cDNA library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES source

Location/Qualifiers
1..434
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CV1-brx-g-10-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CV1"
/note="Vector: pRTD-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CV1
library is a normalized library constructed from rat eye
tissue. For a detailed description of the library from
which this clone was derived, please visit our web site at
rategen.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG TISSUE=rat eye
TAG LIB=UI-R-CV1
TAG_SEQ=CAGCC"

ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 434;
Best Local Similarity 94.7%; Pred. No. 8.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGGCCCCAGCCCCACAA 20
|||||
DB 82 GGGGGCCCCAGCCCCACAA 64

RESULT 21 AA471939/C

LOCUS
DEFINITION
AA471939 461 bp mRNA linear EST 18-JUN-1997
V95H06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
clone IMAGE:873755 5', mRNA sequence.

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AA471939
AA471939.1 GI:2199930
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 461)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE JOURNAL COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:513235

FEATURES source

Location/Qualifiers
1..461
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:873755"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/note="Organ: pooled; Vector: pRTD-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoRI; Site 2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTGCAATCTGAGTGGAGGCGGCGCCCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[GTGATTCGGTACC], digested with Not I and cloned into
the Not I and Eco RI sites of the modified pRTD3 vector.
Library constructed by Bob Barstead."

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 461;
Best Local Similarity 94.7%; Pred. No. 8.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGGCCCCAGCCCCACA 19
|||||
DB 227 CGAGGGCACCAGCCCCACA 209

RESULT 22 W53227/C

LOCUS
DEFINITION
W53227 507 bp mRNA linear EST 03-JUN-1996
nd19a02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone IMAGE:368858 5', mRNA sequence.

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

W53227
W53227.1 GI:1357053
Mus musculus (house mouse)
Mus musculus

REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 507)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE JOURNAL COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project

Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:230290

seq primer: mob.REGA+ET

High quality sequence stop: 358.

FEATURES

source

1. 507

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:368858"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/clone_lib="Soares mouse embryo NbME13.5 14.5"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(GT) primer [5,

TGTTACCAATCTGAAGTGGAGCGCGCGCGAATTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne

State Univ., from 2]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M.Fatima Bonaldo. "

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 507;

Best Local Similarity 94.7%; Pred. No. 8.6e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGCGCCGCGCCGCCACA 19

||||| ||||| ||||| ||||| |||||

Db 202 CGAGCGCCGCGCCGCCACA 184

RESULT 23

AA272464/c

LOCUS

DEFINITION

IMAGE:763438 5', mRNA sequence.

AA272464

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

Waterston,R.

The WashU-HMMI Mouse EST Project

Unpublished (1996)

CONTACT: Marra M/Mouse EST Project

WashU-HMMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:464358

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 491.

FEATURES

source

1. 534

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:763438"

/sex="unknown"

/tissue_type="fetus"

/dev_stage="12.5dpc total fetus"

/lab_host="DH10B"

/clone_lib="Soares mouse 3NME12 5"

/note="Organ: whole fetus; Vector: pT73D-Pac (Pharmacia)

with a modified polylinker; Site 1: Not I; Site 2: Eco RI;

1st strand cDNA was primed with a Not I - oligo(GT) primer

[5, TGTTACCAATCTGAAGTGGAGCGCGCGCGAATTTTTTTTTTTTTTTT

3'], on total mouse RNA [provided by Minoru Ko, Wayne

State Univ.]; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

Library went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match 87.0%; Score 17.4; DB 9; Length 534;

Best Local Similarity 94.7%; Pred. No. 8.7e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGCGCCGCGCCGCCACA 19

||||| ||||| ||||| ||||| |||||

Db 190 CGAGCGCCGCGCCGCCACA 172

RESULT 24

BE395928/c

LOCUS

DEFINITION

601312705F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3659977 5',

mRNA sequence.

BE395928

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

(human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 562)

NIH-MGC http://imgc.ncbi.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ARCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM351 row: i column: 02

High quality sequence stop: 562.

Location/Qualifiers

1. 562

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3659977"

/tissue_type="endometrium, adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_44"

/note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 562;
Best Local Similarity 94.7%; Pred. No. 8.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGCCGACCA 20

Db 74 GAGGCGCCCGCCGACCA 56

RESULT 25

AA458319/c

LOCUS

DEFINITION

AA458319

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:508700

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 450.

Location/Qualifiers

1..588

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:864612"

/sex="male"

/tissue_type="mammary gland"

/dev_stage="4 weeks"

/lab_host="DH10B"

/clone_lib="Soares mammary gland NbMMG"

/note="Organ: mammary gland; Vector: pTV73D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dn) primer [5,

TGTTACCAATCTGAAGTGGAGCGCCGCGATGTTTITTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pTV73 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M. Fatima

Bonaldo."

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 588;
Best Local Similarity 94.7%; Pred. No. 8.8e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGACCA 19

Db 143 CGAGGCGCCCGCCGACCA 125

RESULT 26

AA519009/c

LOCUS

DEFINITION

AA519009

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:523347

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 489.

Location/Qualifiers

1..590

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C3H"

/db_xref="taxon:10090"

/clone="IMAGE:903283"

/cell_line="C2C12"

/lab_host="DH10B"

/clone_lib="Barstead mouse myotubes MPLRB5"

/note="Vector: pTV73D-Pac (Pharmacia) with a modified

polylinker; Site 1: EcoRI; Site 2: NotI; 1st strand cDNA

was primed with a Not I - oligo(dn) primer [5,

TGTTACCAATCTGAAGTGGAGCGCCGCGATGTTTITTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

[AATTCGATCCTTG], digested with Not I and cloned into the

Not I and Eco RI sites of the modified pTV73 vector.

Library constructed by Bob Barstead. The C2C12 cell line

(available from ATCC, catalog # CRL-1772) differentiates

rapidly, forming contractile myotubes and producing

characteristic muscle proteins."

ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 590;
Best Local Similarity 94.7%; Pred. No. 8.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCGACCA 19

Db 208 CGAGGCGCCCGCCGACCA 190

RESULT 27
 BI838149
 LOCUS
 DEFINITION
 603083654P1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5222917 5',
 mRNA sequence.
 ACCESSION
 VERSION
 BI838149.1 GI:15949699
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 781)
 NIH-MGC http://mgc.nci.nih.gov/.
 TITLE
 JOURNAL
 COMMENT
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM11560 row: k column: 14
 High quality sequence stop: 342.
 Location/Qualifiers
 1..781
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5222917"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC 120"
 /note="Organ: pooled pancreas and spleen; Vector:
 pCMV-SPORT5; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
 source anonymous pool of spleen and pancreas from 28 yo
 male. Library is oligo-dT primed and directionally cloned
 (EcoRV site is destroyed upon cloning). Average insert
 size 1.5 kb, insert size range 1-2.5 kb. Library is
 normalized and enriched for full-length clones and was
 constructed by C. Gruber (Invitrogen). Research Genetics
 tracking code 025. Note: this is a NIH_MGC Library."
 ORIGIN
 Query Match 87.0%; Score 17.4; DB 12; Length 781;
 Best Local Similarity 94.7%; Pred. No. 9.4e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 GAGGCGCCAGCCACAA 20
 Db 695 GAGGCGCTCAGCCACAA 713
 RESULT 28
 BF260146
 LOCUS
 DEFINITION
 HVSMF0021D17f Hordeum vulgare seedling root EST library HVCDNA0007
 (stiolated and unstressed) Hordeum vulgare subsp. vulgare cDNA
 clone HVSMF0021D17f, mRNA sequence.
 ACCESSION
 VERSION
 BF260146.3 GI:16316189
 KEYWORDS
 SOURCE
 ORGANISM
 Hordeum vulgare subsp. vulgare
 Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooideae; Triticeae; Hordeum.
 1 (bases 1 to 893)
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D.,
 Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W.,
 Fenton, R.D., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex unstressed seedling root cDNA library
 Unpublished (2001)
 On Nov 16, 2000 this sequence version replaced gi:13120492.
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 366
 Seq primer: AATAACCTCACTAAAGG
 High quality sequence stop: 753.
 Location/Qualifiers
 1..893
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVSMF0021D17f"
 /tissue_type="Seedling root"
 /lab_host="HJC121"
 /clone_lib="Hordeum vulgare seedling root EST library
 HVCDNA0007 (Etiolated and unstressed)"
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper
 with water, nystatin and cefotaxime in covered
 crystallization dishes. Five-day old seedling roots were
 then harvested, total RNA was prepared, poly(A) RNA was
 purified, one primary unamplified cDNA library was made,
 and 1 million pfu were in vivo excised to give pBluescript
 SK(-) cDNA phagemids. These steps were performed in the TJ
 Close laboratory at the University of California,
 Riverside (Choi, Close, Fenton). Phagemids were plated and
 picked at the Clemson University Genomics Institute (CUGI)
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
 Oates, Rambo, Main). The sequence has been trimmed to
 remove vector sequence and contains a minimum of 100 bases
 of phred value 20 or above. For more details on library
 preparation and sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders/Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

TITLE
 JOURNAL
 COMMENT
 ORIGIN
 Query Match 87.0%; Score 17.4; DB 10; Length 893;
 Best Local Similarity 94.7%; Pred. No. 9.7e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 CGAGCGCCAGCCACCA 19
 Db 691 CGAGCGCCAGCCACCA 709
 RESULT 29
 BZ569703
 LOCUS
 DEFINITION
 pacs2-164_979.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
 pacs2-164_979, genomic survey sequence.
 ACCESSION
 VERSION
 BZ569703.1 GI:27204692
 KEYWORDS
 SOURCE
 Pseudomonas aeruginosa

PRIMERS
 Sequencing: TJ
 LIBRARY
 Vector : pBACE3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.

Search completed: June 20, 2004, 14:13:08
Job time : 962.603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 22.1164 Seconds
(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-11

Perfect score: 20

Sequence: 1 cgaggcgccagcccccacaa 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| 1 | 19 | 95.0 | 4403765 | 3 | US-09-103-840A-2 |
| 2 | 19 | 95.0 | 4411529 | 3 | US-09-103-840A-1 |
| 3 | 16 | 80.0 | 3845 | 2 | US-08-220-240A-1 |
| 4 | 15.8 | 79.0 | 477 | 4 | US-09-621-976-18886 |
| 5 | 15.8 | 79.0 | 947 | 4 | US-09-347-878-49 |
| 6 | 15.8 | 79.0 | 981 | 3 | US-08-815-225-1 |
| 7 | 15.8 | 79.0 | 4380 | 4 | US-08-955-565A-3 |
| 8 | 15.8 | 79.0 | 13857 | 4 | US-09-620-312D-75 |
| 9 | 15.8 | 79.0 | 4403765 | 3 | US-09-103-840A-2 |
| 10 | 15.8 | 79.0 | 4411529 | 3 | US-09-103-840A-1 |
| 11 | 15.4 | 77.0 | 216 | 4 | US-09-252-991A-1061 |
| 12 | 15.4 | 77.0 | 432 | 4 | US-09-252-991A-3530 |
| 13 | 15.4 | 77.0 | 504 | 4 | US-09-252-991A-3564 |
| 14 | 15.4 | 77.0 | 636 | 4 | US-09-252-991A-7251 |
| 15 | 15.4 | 77.0 | 708 | 4 | US-09-252-991A-13744 |
| 16 | 15.4 | 77.0 | 804 | 4 | US-09-252-991A-13605 |
| 17 | 15.4 | 77.0 | 807 | 4 | US-09-252-991A-13398 |
| 18 | 15.4 | 77.0 | 858 | 4 | US-09-033-556-8 |
| 19 | 15.4 | 77.0 | 858 | 4 | US-09-474-699-12 |
| 20 | 15.4 | 77.0 | 858 | 4 | US-09-151-376-8 |
| 21 | 15.4 | 77.0 | 963 | 4 | US-09-252-991A-7371 |
| 22 | 15.4 | 77.0 | 984 | 4 | US-09-252-991A-13494 |
| 23 | 15.4 | 77.0 | 1002 | 4 | US-09-252-991A-3539 |
| 24 | 15.4 | 77.0 | 1143 | 4 | US-09-252-991A-949 |
| 25 | 15.4 | 77.0 | 1350 | 4 | US-09-252-991A-1098 |
| 26 | 15.4 | 77.0 | 1536 | 4 | US-09-489-039A-1997 |
| 27 | 15.4 | 77.0 | 1644 | 4 | US-09-252-991A-3548 |

| | | | | | |
|---|---------------------|--------|------|------|--------------------|
| 4 | US-09-252-991A-1134 | 1791 | 77.0 | 15.4 | Sequence 1134, Ap |
| 4 | US-09-301-978C-1 | 2221 | 77.0 | 15.4 | Sequence 1, Appli |
| 4 | US-09-220-132-188 | 2274 | 77.0 | 15.4 | Sequence 188, Appl |
| 3 | US-09-252-991A-3523 | 2832 | 77.0 | 15.4 | Sequence 3523, Ap |
| 3 | US-09-339-964-1 | 4743 | 77.0 | 15.4 | Sequence 1, Appli |
| 3 | US-08-922-635-21 | 15202 | 77.0 | 15.4 | Sequence 21, Appl |
| 4 | US-09-740-027-3 | 24707 | 77.0 | 15.4 | Sequence 3, Appli |
| 4 | US-09-833-381-649 | 412 | 76.0 | 15.2 | Sequence 649, App |
| 4 | US-09-489-039A-246 | 1038 | 76.0 | 15.2 | Sequence 246, App |
| 4 | US-09-252-991A-892 | 1257 | 76.0 | 15.2 | Sequence 892, App |
| 4 | US-09-252-991A-1139 | 1575 | 76.0 | 15.2 | Sequence 1139, Ap |
| 4 | US-09-252-991A-944 | 2859 | 76.0 | 15.2 | Sequence 944, App |
| 3 | US-09-318-448-20 | 3228 | 76.0 | 15.2 | Sequence 20, Appl |
| 5 | PCT-US95-08354A-1 | 4016 | 76.0 | 15.2 | Sequence 1, Appli |
| 4 | US-09-147-236-1 | 16836 | 76.0 | 15.2 | Sequence 1, Appli |
| 4 | US-09-147-236-10 | 16836 | 76.0 | 15.2 | Sequence 10, Appl |
| 4 | US-09-522-474-1 | 16836 | 76.0 | 15.2 | Sequence 1, Appli |
| 4 | US-09-522-474-10 | 16836 | 76.0 | 15.2 | Sequence 10, Appl |
| 4 | US-09-489-039A-4030 | 738 | 75.0 | 15 | Sequence 4030, Ap |
| 4 | US-09-252-991A-1516 | 795 | 75.0 | 15 | Sequence 1516, Ap |
| 4 | US-09-252-991A-1557 | 831 | 75.0 | 15 | Sequence 1557, Ap |
| 4 | US-09-252-991A-1662 | 1821 | 75.0 | 15 | Sequence 1662, Ap |
| 4 | US-09-489-039A-2928 | 1878 | 75.0 | 15 | Sequence 2928, Ap |
| 4 | US-09-252-991A-1466 | 2346 | 75.0 | 15 | Sequence 1466, Ap |
| 4 | US-09-489-039A-3025 | 405 | 74.0 | 14.8 | Sequence 3025, Ap |
| 4 | US-09-976-594-305 | 452 | 74.0 | 14.8 | Sequence 305, App |
| 4 | US-09-252-991A-6706 | 456 | 74.0 | 14.8 | Sequence 6706, Ap |
| 4 | US-09-621-976-817 | 507 | 74.0 | 14.8 | Sequence 817, App |
| 4 | US-09-833-381-1619 | 512 | 74.0 | 14.8 | Sequence 1619, Ap |
| 4 | US-09-976-594-973 | 518 | 74.0 | 14.8 | Sequence 973, App |
| 4 | US-09-621-976-3035 | 522 | 74.0 | 14.8 | Sequence 3035, Ap |
| 4 | US-09-621-976-929 | 547 | 74.0 | 14.8 | Sequence 929, App |
| 3 | US-09-385-982-143 | 620 | 74.0 | 14.8 | Sequence 143, App |
| 4 | US-09-620-312D-506 | 649 | 74.0 | 14.8 | Sequence 506, App |
| 4 | US-09-252-991A-8721 | 678 | 74.0 | 14.8 | Sequence 8721, Ap |
| 4 | US-09-621-976-1969 | 708 | 74.0 | 14.8 | Sequence 1969, Ap |
| 4 | US-09-599-360B-44 | 715 | 74.0 | 14.8 | Sequence 44, Appl |
| 4 | US-09-252-991A-8940 | 822 | 74.0 | 14.8 | Sequence 8940, Ap |
| 4 | US-09-252-991A-6673 | 834 | 74.0 | 14.8 | Sequence 6673, Ap |
| 4 | US-09-252-991A-8823 | 1110 | 74.0 | 14.8 | Sequence 8823, Ap |
| 4 | US-09-833-381-1621 | 1227 | 74.0 | 14.8 | Sequence 1621, Ap |
| 2 | US-08-851-088-7 | 1275 | 74.0 | 14.8 | Sequence 251, Appl |
| 3 | US-08-851-089-1 | 1362 | 74.0 | 14.8 | Sequence 1, Appli |
| 3 | US-09-833-381-1620 | 1407 | 74.0 | 14.8 | Sequence 1620, Ap |
| 3 | US-09-344-001-1 | 1461 | 74.0 | 14.8 | Sequence 1, Appli |
| 2 | US-08-039-297B-1 | 1645 | 74.0 | 14.8 | Sequence 1, Appli |
| 4 | US-09-252-991A-9061 | 1734 | 74.0 | 14.8 | Sequence 9061, Ap |
| 4 | US-09-206-166-3 | 1775 | 74.0 | 14.8 | Sequence 3, Appli |
| 4 | US-09-252-991A-281 | 1785 | 74.0 | 14.8 | Sequence 281, App |
| 1 | US-07-832-855-1 | 1905 | 74.0 | 14.8 | Sequence 1, Appli |
| 4 | US-08-176-320-1 | 1905 | 74.0 | 14.8 | Sequence 1, Appli |
| 3 | US-09-165-042-2 | 1976 | 74.0 | 14.8 | Sequence 2, Appli |
| 4 | US-09-252-991A-9144 | 2154 | 74.0 | 14.8 | Sequence 9144, Ap |
| 4 | US-09-252-991A-6641 | 2163 | 74.0 | 14.8 | Sequence 6641, Ap |
| 4 | US-09-613-444-1 | 2174 | 74.0 | 14.8 | Sequence 1, Appli |
| 4 | US-09-620-312D-909 | 2268 | 74.0 | 14.8 | Sequence 909, App |
| 4 | US-09-280-116-202 | 2276 | 74.0 | 14.8 | Sequence 202, App |
| 4 | US-09-206-166-9 | 2672 | 74.0 | 14.8 | Sequence 9, Appli |
| 4 | US-09-232-279-1 | 2799 | 74.0 | 14.8 | Sequence 1, Appli |
| 1 | US-08-398-008A-1 | 2823 | 74.0 | 14.8 | Sequence 1, Appli |
| 2 | US-08-893-333-1 | 2823 | 74.0 | 14.8 | Sequence 1, Appli |
| 2 | US-08-682-847-1 | 3382 | 74.0 | 14.8 | Sequence 1, Appli |
| 1 | US-08-035-558-1 | 3519 | 74.0 | 14.8 | Sequence 1, Appli |
| 3 | US-08-851-089-12 | 4144 | 74.0 | 14.8 | Sequence 12, Appl |
| 3 | US-08-851-089-13 | 4144 | 74.0 | 14.8 | Sequence 13, Appl |
| 2 | US-08-206-176-7 | 10807 | 74.0 | 14.8 | Sequence 7, Appli |
| 2 | US-08-756-506-5 | 10807 | 74.0 | 14.8 | Sequence 5, Appli |
| 4 | US-09-816-095-3 | 99916 | 74.0 | 14.8 | Sequence 3, Appli |
| 4 | US-09-754-250-3 | 111282 | 74.0 | 14.8 | Sequence 32, Appl |
| 4 | US-09-497-855A-32 | 118067 | 74.0 | 14.8 | Sequence 3, Appli |
| 4 | US-09-804-471A-3 | 174493 | 74.0 | 14.8 | Sequence 3, Appli |
| 4 | US-10-238-709-3 | 174493 | 74.0 | 14.8 | Sequence 3, Appli |


```
C 247 13.8 69.0 439 4 US-09-621-976-9244
C 248 13.8 69.0 452 4 US-09-621-976-9111
C 249 13.8 69.0 460 4 US-09-621-976-3265
C 250 13.8 69.0 488 4 US-09-280-116-109
C 251 13.8 69.0 522 4 US-09-621-976-8994
C 252 13.8 69.0 591 2 US-08-481-814A-4
C 253 13.8 69.0 593 4 US-09-517-431E-10
C 254 13.8 69.0 618 4 US-09-489-039A-2443
C 255 13.8 69.0 680 4 US-09-833-381-976
C 256 13.8 69.0 699 3 US-08-998-416-930
C 257 13.8 69.0 764 3 US-09-188-930-263
C 258 13.8 69.0 764 4 US-09-724-864-66
C 259 13.8 69.0 764 4 US-09-312-283C-263
C 260 13.8 69.0 766 3 US-09-188-930-38
C 261 13.8 69.0 766 4 US-09-312-283C-38
C 262 13.8 69.0 771 4 US-09-489-039A-4829
C 263 13.8 69.0 789 4 US-09-252-991A-12447
C 264 13.8 69.0 798 4 US-09-252-991A-2808
C 265 13.8 69.0 877 4 US-09-204-865-7
C 266 13.8 69.0 900 4 US-09-489-039A-5980
C 267 13.8 69.0 960 4 US-09-489-039A-308
C 268 13.8 69.0 971 4 US-09-023-655-646
C 269 13.8 69.0 972 4 US-09-252-991A-9101
C 270 13.8 69.0 1008 4 US-09-252-991A-10020
C 271 13.8 69.0 1029 4 US-09-252-991A-15460
C 272 13.8 69.0 1044 4 US-09-252-991A-2435
C 273 13.8 69.0 1062 4 US-09-244-111-1
C 274 13.8 69.0 1098 4 US-09-252-991A-15400
C 275 13.8 69.0 1100 2 US-08-481-814A-5
C 276 13.8 69.0 1107 4 US-09-173-300-50
C 277 13.8 69.0 1134 4 US-09-252-991A-7137
C 278 13.8 69.0 1170 4 US-09-252-991A-7929
C 279 13.8 69.0 1194 4 US-09-252-991A-11352
C 280 13.8 69.0 1248 4 US-09-252-991A-7361
C 281 13.8 69.0 1294 2 US-09-665-647-4
C 282 13.8 69.0 1308 4 US-09-489-039A-2811
C 283 13.8 69.0 1314 4 US-09-242-737-1
C 284 13.8 69.0 1363 1 US-08-776-088-21
C 285 13.8 69.0 1363 5 PCT-US95-09145A-21
C 286 13.8 69.0 1377 4 US-09-252-991A-2251
C 287 13.8 69.0 1404 4 US-09-252-991A-7623
C 288 13.8 69.0 1404 4 US-09-489-039A-2728
C 289 13.8 69.0 1434 4 US-09-489-039A-2766
C 290 13.8 69.0 1461 4 US-09-894-844-84
C 291 13.8 69.0 1488 4 US-09-252-991A-15378
C 292 13.8 69.0 1500 4 US-09-252-991A-11312
C 293 13.8 69.0 1509 4 US-09-489-039A-6014
C 294 13.8 69.0 1542 3 US-09-088-271A-13
C 295 13.8 69.0 1551 4 US-09-489-039A-4788
C 296 13.8 69.0 1663 4 US-09-312-283C-370
C 297 13.8 69.0 1663 4 US-09-312-283C-416
C 298 13.8 69.0 1829 4 US-09-620-312D-274
C 299 13.8 69.0 1830 4 US-09-252-991A-3164
C 300 13.8 69.0 1842 4 US-09-252-991A-15508
```

ALIGNMENTS

```
RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 95.0%; Score 19; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 GAGGCGCCGAGCCCAACA 20
Db 2599945 GAGGCGCCGAGCCCAACA 2599963
```

```
RESULT 2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match 95.0%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 GAGGCGCCGAGCCCAACA 20
Db 2604135 GAGGCGCCGAGCCCAACA 2604153
```

```
RESULT 3
US-08-220-240A-4/c
; Sequence 4, Application US/08220240A
; Patent No. 5955291
; GENERAL INFORMATION:
; APPLICANT: Alitalo, Kari
; APPLICANT: Matikainen, Marja-Terttu
; APPLICANT: Partanen, Juha
; APPLICANT: Makela, Tomi
; APPLICANT: Korhonen, Jaana
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING TIE RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASE AND USES THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
```

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220,240A
; FILING DATE: 29-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI93/00006
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/817,800
; FILING DATE: 09-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,453
; FILING DATE: 15-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 29151/31958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 37..3450
; US-08-220-240A-4

Query Match 80.0%; Score 16; DB 2; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCGCCACAA 20
DB 693 GCGCCAGCCGCCACAA 678

RESULT 4
US-09-621-976-18886/c
; Sequence 18886, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. Y.
; APPLICANT: Giordano, J. Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18886
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18886

Query Match 79.0%; Score 15.8; DB 4; Length 477;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCGCCACAA 20
DB 292 GAGGCTCCCTGCCGCCACAA 274

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220,240A
; FILING DATE: 29-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI93/00006
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/817,800
; FILING DATE: 09-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/167,453
; FILING DATE: 15-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 29151/31958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 37..3450
; US-08-220-240A-4

Query Match 80.0%; Score 16; DB 2; Length 3845;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCGCCACAA 20
DB 693 GCGCCAGCCGCCACAA 678

RESULT 4
US-09-621-976-18886/c
; Sequence 18886, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. Y.
; APPLICANT: Giordano, J. Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18886
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18886

Query Match 79.0%; Score 15.8; DB 4; Length 477;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCGCCACAA 20
DB 292 GAGGCTCCCTGCCGCCACAA 274
```

```
RESULT 5
US-09-347-878-49/c
; Sequence 49, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 49
; LENGTH: 947
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)..(801)
; FEATURE:
; OTHER INFORMATION: Human short-chain alcohol dehydrogenase cDNA
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: U73514/GenBank
US-09-347-878-49
```

```
Query Match 79.0%; Score 15.8; DB 4; Length 947;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCAGCCGCCACAA 20
DB 122 GAGGCTCCCTGCCGCCACAA 104
```

```
RESULT 6
US-08-815-225-1/c
; Sequence 1, Application US/08815225
; Patent No. 6268479
; GENERAL INFORMATION:
; APPLICANT: Stern, David M.
; APPLICANT: Yan, Shi Du
; TITLE OF INVENTION: AN INTRACELLULAR AMYLOID-BETA PEPTIDE
; TITLE OF INVENTION: BINDING (BRAB) POLYPEPTIDE AND METHODS FOR INHIBITING
; TITLE OF INVENTION: NEURODEGENERATIVE CONDITIONS.
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/815,225
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/55209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 981 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..801
US-08-815-225-1

Query Match          79.0%; Score 15.8; DB 3; Length 981;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 GAGGCGCCCGCCGCCACAA 20
Db      125 GAGGCTCCCTGCCGCCACAA 107

RESULT 7
US-08-955-565A-3/c
; Sequence 3, Application US/08955565A
; Patent No. 6331388
; GENERAL INFORMATION:
; APPLICANT: Malkovsky, Miroslav
; TITLE OF INVENTION: Immune Response Enhancer Therapy
; FILE REFERENCE: WARF-02625
; CURRENT APPLICATION NUMBER: US/08/955,565A
; CURRENT FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4380
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-08-955-565A-3

Query Match          79.0%; Score 15.8; DB 4; Length 4380;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 GAGGCGCCCGCCGCCACAA 20
Db      1929 GAGGAGCCCGGCCGCCACAA 1911

RESULT 8
US-09-620-312D-75
; Sequence 75, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John, Tillinghast
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
```

```
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 75
; LENGTH: 13857
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(13857)
US-09-620-312D-75

Query Match          79.0%; Score 15.8; DB 4; Length 13857;
Best Local Similarity 89.5%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CGAGGCGCCCGCCGCCACAA 19
Db      4956 CGAGACCCCGCCGCCACAA 4974

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          79.0%; Score 15.8; DB 3; Length 4403765;
Best Local Similarity 89.5%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 GAGGCGCCCGCCGCCACAA 20
Db      531725 GAGGAGCCCGCCGCCACAA 531707

RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          79.0%; Score 15.8; DB 3; Length 4411529;
Best Local Similarity 89.5%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGGCGCCCGAGCCCAACAA 20
   |||||
Db 530283 GAGGAGCCCGCCCAACAA 530265

RESULT 11
US-09-252-991A-1061/c
; Sequence 1061, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1061
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1061

Query Match          77.0%; Score 15.4; DB 4; Length 216;
Best Local Similarity 94.1%; Pred. No. 3.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGAGCCCA 17
   |||||
Db 207 CGTGGCGCCCGAGCCCA 191

RESULT 12
US-09-252-991A-3530/c
; Sequence 3530, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3530
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3530

Query Match          77.0%; Score 15.4; DB 4; Length 432;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

US-09-252-991A-7251/c
; Sequence 7251, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7251
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7251

Query Match          77.0%; Score 15.4; DB 4; Length 636;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGAGCCCA 17
   |||||
Db 131 CGTGGCGCCCGAGCCCA 147

RESULT 14
US-09-252-991A-7251/c
; Sequence 7251, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7251
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7251

Query Match          77.0%; Score 15.4; DB 4; Length 636;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGAGCCCA 17
   |||||
Db 168 CAAGGCGCCCGAGCCCA 152

RESULT 15
US-09-252-991A-13744
; Sequence 13744, Application US/09252991A
; Patent No. 6551795
```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13744
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13744

Query Match 77.0%; Score 15.4; DB 4; Length 708;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
   |||||
Db 460 CGCGCGCCCGCCCA 476

RESULT 16
US-09-252-991A-13605/c
; Sequence 13605, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13605
; LENGTH: 804
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13605

Query Match 77.0%; Score 15.4; DB 4; Length 804;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
   |||||
Db 286 CGCGCGCCCGCCCA 270

RESULT 17
US-09-252-991A-13398/c
; Sequence 13398, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13398
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13398

Query Match 77.0%; Score 15.4; DB 4; Length 807;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGGCGCCCGCCCA 17
   |||||
Db 717 CGCGCGCCCGCCCA 701

RESULT 18
US-09-033-556-8
; Sequence 8, Application US/09033556
; Patent No. 6432700
; GENERAL INFORMATION:
; APPLICANT: Henderson, Daniel R.
; TITLE OF INVENTION: ADENOVIRUS VECTORS CONTAINING
; TITLE OF INVENTION: HETEROLOGOUS TRANSCRIPTION REGULATORY ELEMENTS AND METHODS
; TITLE OF INVENTION: OF USING SAME
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,556
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Catherine, Polizzi M
; REGISTRATION NUMBER: 40,130
; REFERENCE/DOCKET NUMBER: 34802-20010.00
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-033-556-8

Query Match 77.0%; Score 15.4; DB 4; Length 858;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCGCCCGCCCGCCCA 20
   |||||
Db 20 GGCGCCCGCCCGCCCA 36

RESULT 19
US-09-474-699-12
```

; Sequence 12, Application US/09474699
; Patent No. 6495130
; GENERAL INFORMATION:
; APPLICANT: Henderson, Daniel R.
; APPLICANT: Yu, De Chao
; TITLE OF INVENTION: TARGET CELL-SPECIFIC ADENOVIRAL VECTORS
; TITLE OF INVENTION: CONTAINING E3 AND METHODS OF USE THEREOF
; FILE REFERENCE: 348022001300
; CURRENT APPLICATION NUMBER: US/09/474,699
; CURRENT FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/114,262
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-474-699-12

Query Match 77.0%; Score 15.4; DB 4; Length 858;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCGCCCGCCCAAA 20
DB 20 GGCGCCCGCCCGCAA 36

RESULT 20

US-09-151-376-8
; Sequence 8, Application US/09151376
; Patent No. 6676935
; GENERAL INFORMATION:
; APPLICANT: Henderson, D.R.
; APPLICANT: Schuur, E.R.
; TITLE OF INVENTION: TISSUE SPECIFIC VIRAL VECTORS
; FILE REFERENCE: 34802200221
; CURRENT APPLICATION NUMBER: US/09/151,376
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: 08/669,753
; EARLIER FILING DATE: 1996-06-26
; EARLIER APPLICATION NUMBER: 08/495,034
; EARLIER FILING DATE: 1995-06-27
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-151-376-8

Query Match 77.0%; Score 15.4; DB 4; Length 858;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCGCCCGCCCAAA 20
DB 20 GGCGCCCGCCCGCAA 36

RESULT 21

US-09-252-991A-7371
; Sequence 7371, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7371
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7371

Query Match 77.0%; Score 15.4; DB 4; Length 963;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGCGCCCGCCCA 17
DB 14 CAAGCGCCCGCCCA 30

RESULT 22

US-09-252-991A-13494/c
; Sequence 13494, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13494
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13494

Query Match 77.0%; Score 15.4; DB 4; Length 984;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGAGCGCCCGCCCA 17
DB 128 CGCGCCCGCCCGCCCA 112

RESULT 23

US-09-252-991A-3539/c
; Sequence 3539, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3539
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3539

Query Match 77.0%; Score 15.4; DB 4; Length 1002;

Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGGCCCGAGCCCA 17
|||
Db 343 CGTGGCGCCCGAGCCCA 327

RESULT 24
US-09-252-991A-949
; Sequence 949, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 949
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-949

Query Match 77.0%; Score 15.4; DB 4; Length 1143;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGGCCCGAGCCCA 17
|||
Db 1059 CGTGGCGCCCGAGCCCA 1075

RESULT 25
US-09-252-991A-1098/c
; Sequence 1098, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1098
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1098

Query Match 77.0%; Score 15.4; DB 4; Length 1350;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGGCCCGAGCCCA 17
|||
Db 785 CGTGGCGCCCGAGCCCA 769

RESULT 26
US-09-489-039A-1997/c
; Sequence 1997, Application US/09489039A

Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1997
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1997

Query Match 77.0%; Score 15.4; DB 4; Length 1536;
Best Local Similarity 94.1%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGGCGCCCGAGCCCA 18
|||
Db 653 GAGGCGCCCGAGCCCA 637

RESULT 27
US-09-252-991A-3548
; Sequence 3548, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3548
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3548

Query Match 77.0%; Score 15.4; DB 4; Length 1644;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGGCCCGAGCCCA 17
|||
Db 754 CGTGGCGCCCGAGCCCA 770

RESULT 28
US-09-252-991A-1134/c
; Sequence 1134, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 1134
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1134

Query Match      77.0%; Score 15.4; DB 4; Length 1791;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CGAGGCGCCCGAGCCCA 17
Db      736 CGTGGCGCCCGAGCCCA 720

; ORGANISM: Homo sapiens
US-09-220-132-188

Query Match      77.0%; Score 15.4; DB 4; Length 2274;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GAGGCGCCCGAGCCCGAC 18
Db      1584 GAGGCGCCCGAGCCCGAC 1568

Search completed: June 20, 2004, 11:45:36
Job time : 39.2414 secs
```

```
RESULT 29
US-09-301-978C-1
; Sequence 1, Application US/09301978C
; Patent No. 6392015
; GENERAL INFORMATION:
; APPLICANT: Panganiban, Antonito
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
; TITLE OF INVENTION: Interaction with U Binding Protein (UBP)
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/09/301,978C
; CURRENT FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2221
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1514)
; OTHER INFORMATION: n = any nucleotide.
; NAME/KEY: unsure
; LOCATION: (2066)
; OTHER INFORMATION: n = any nucleotide.
US-09-301-978C-1

Query Match      77.0%; Score 15.4; DB 4; Length 2221;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GAGGCGCCCGAGCCCGAC 18
Db      660 GAGGCGCCCGAGCCCGAC 676

RESULT 30
US-09-220-132-188/c
; Sequence 188, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shivjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 2274
; TYPE: DNA
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 113.122 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-11
Perfect score: 20
Sequence: 1 CGAGGCGCCAGCCCAACA 20

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Published Applications NA:**

| Result No. | Query Match | Length | ID | Description |
|------------|-------------|--------|----------|----------------------|
| C 1 | 16.8 | 84.0 | 993 14 | US-10-001-857-91 |
| C 2 | 16.8 | 84.0 | 1002 14 | US-10-001-873-6 |
| C 3 | 16.8 | 84.0 | 1094 13 | US-10-027-632-9289 |
| C 4 | 16.8 | 84.0 | 1094 13 | US-10-027-632-9290 |
| C 5 | 16.8 | 84.0 | 1094 13 | US-10-027-632-9291 |
| C 6 | 16.8 | 84.0 | 1094 16 | US-10-027-632-9289 |
| C 7 | 16.8 | 84.0 | 1094 16 | US-10-027-632-9290 |
| C 8 | 16.8 | 84.0 | 1094 16 | US-10-027-632-9291 |
| C 9 | 16.8 | 84.0 | 1830 13 | US-10-282-122A-40410 |
| C 10 | 16.8 | 84.0 | 9851 15 | US-10-017-161-1677 |
| C 11 | 16.8 | 84.0 | 9851 16 | US-10-292-788-1337 |
| C 12 | 16.8 | 84.0 | 21501 15 | US-10-006-430-11 |
| C 13 | 16.8 | 84.0 | 28000 15 | US-10-091-625-11 |
| C 14 | 16.8 | 84.0 | 28000 15 | US-10-096-399A-11 |
| C 15 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 16 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 17 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 18 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 19 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 20 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 21 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 22 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 23 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 24 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 25 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 26 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 27 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 28 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 29 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 30 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 31 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 32 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 33 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 34 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 35 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 36 | 16.8 | 84.0 | 28000 16 | US-10-332-859-324 |
| C 37 | 15.8 | 79.0 | 245 13 | US-10-424-599-40951 |
| C 38 | 15.8 | 79.0 | 268 13 | US-10-085-783A-11265 |
| C 39 | 15.8 | 79.0 | 268 16 | US-10-242-535A-11265 |
| C 40 | 15.8 | 79.0 | 340 13 | US-10-424-599-83608 |
| C 41 | 15.8 | 79.0 | 341 13 | US-10-424-599-58186 |
| C 42 | 15.8 | 79.0 | 461 10 | US-09-918-995-28160 |
| C 43 | 15.8 | 79.0 | 492 10 | US-09-918-995-13759 |
| C 44 | 15.8 | 79.0 | 786 9 | US-09-931-186-1 |
| C 45 | 15.8 | 79.0 | 786 9 | US-09-931-186-3 |
| C 46 | 15.8 | 79.0 | 786 9 | US-09-931-186-5 |
| C 47 | 15.8 | 79.0 | 786 9 | US-09-931-186-7 |
| C 48 | 15.8 | 79.0 | 786 9 | US-09-931-186-19 |
| C 49 | 15.8 | 79.0 | 786 9 | US-09-931-186-22 |
| C 50 | 15.8 | 79.0 | 874 13 | US-10-027-632-142002 |
| C 51 | 15.8 | 79.0 | 874 13 | US-10-027-632-169596 |
| C 52 | 15.8 | 79.0 | 874 16 | US-10-027-63 |

| | | | | | | |
|-------|------|------|-----|----|----------------------|--------------------|
| C 234 | 14.8 | 74.0 | 253 | 16 | US-10-227-577-1588 | Sequence 1588, Ap |
| C 235 | 14.8 | 74.0 | 297 | 13 | US-10-424-599-102542 | Sequence 102542, A |
| C 236 | 14.8 | 74.0 | 332 | 9 | US-09-867-701-8076 | Sequence 8076, Ap |
| C 237 | 14.8 | 74.0 | 334 | 15 | US-10-029-386-16172 | Sequence 16172, A |
| C 238 | 14.8 | 74.0 | 353 | 9 | US-09-867-701-7715 | Sequence 7715, Ap |
| C 239 | 14.8 | 74.0 | 391 | 9 | US-09-867-701-10054 | Sequence 10054, A |
| C 240 | 14.8 | 74.0 | 426 | 9 | US-09-922-217-5883 | Sequence 583, App |
| C 241 | 14.8 | 74.0 | 426 | 9 | US-09-833-363-583 | Sequence 583, App |
| C 242 | 14.8 | 74.0 | 426 | 14 | US-10-025-380-583 | Sequence 583, App |
| C 243 | 14.8 | 74.0 | 438 | 9 | US-09-864-761-4805 | Sequence 4805, App |
| C 244 | 14.8 | 74.0 | 467 | 13 | US-10-027-632-27630 | Sequence 27630, A |
| C 245 | 14.8 | 74.0 | 467 | 16 | US-10-027-632-27630 | Sequence 27630, A |
| C 246 | 14.8 | 74.0 | 482 | 13 | US-10-424-599-140369 | Sequence 140369, A |
| C 247 | 14.8 | 74.0 | 501 | 10 | US-09-918-995-19115 | Sequence 19115, A |
| C 248 | 14.8 | 74.0 | 502 | 16 | US-10-242-355-700 | Sequence 700, App |
| C 249 | 14.8 | 74.0 | 511 | 13 | US-10-027-632-127008 | Sequence 127008, A |
| C 250 | 14.8 | 74.0 | 511 | 16 | US-10-027-632-127008 | Sequence 127008, A |
| C 251 | 14.8 | 74.0 | 512 | 9 | US-09-833-381-1619 | Sequence 1619, App |
| C 252 | 14.8 | 74.0 | 524 | 13 | US-10-027-632-107300 | Sequence 107300, A |
| C 253 | 14.8 | 74.0 | 524 | 16 | US-10-027-632-107300 | Sequence 107300, A |
| C 254 | 14.8 | 74.0 | 532 | 13 | US-10-027-632-136959 | Sequence 136959, A |
| C 255 | 14.8 | 74.0 | 532 | 16 | US-10-027-632-136959 | Sequence 136959, A |
| C 256 | 14.8 | 74.0 | 565 | 9 | US-09-864-761-7809 | Sequence 7809, App |
| C 257 | 14.8 | 74.0 | 576 | 10 | US-09-814-353-18852 | Sequence 18852, A |
| C 258 | 14.8 | 74.0 | 583 | 15 | US-10-029-386-2472 | Sequence 2472, App |
| C 259 | 14.8 | 74.0 | 598 | 9 | US-09-895-828-323 | Sequence 323, App |
| C 260 | 14.8 | 74.0 | 598 | 15 | US-10-114-666-323 | Sequence 323, App |
| C 261 | 14.8 | 74.0 | 609 | 15 | US-10-198-846-6143 | Sequence 6143, App |
| C 262 | 14.8 | 74.0 | 609 | 15 | US-10-156-761-4684 | Sequence 4684, App |
| C 263 | 14.8 | 74.0 | 620 | 10 | US-09-871-161-143 | Sequence 143, App |
| C 264 | 14.8 | 74.0 | 624 | 9 | US-09-893-737-159 | Sequence 159, App |
| C 265 | 14.8 | 74.0 | 638 | 9 | US-09-777-564-651 | Sequence 651, App |
| C 266 | 14.8 | 74.0 | 638 | 15 | US-10-015-219-651 | Sequence 651, App |
| C 267 | 14.8 | 74.0 | 640 | 12 | US-09-978-360A-215 | Sequence 215, App |
| C 268 | 14.8 | 74.0 | 649 | 15 | US-10-037-270-506 | Sequence 506, App |
| C 269 | 14.8 | 74.0 | 649 | 15 | US-10-106-698-3696 | Sequence 3696, App |
| C 270 | 14.8 | 74.0 | 649 | 16 | US-10-117-722-506 | Sequence 506, App |
| C 271 | 14.8 | 74.0 | 650 | 13 | US-10-320-120-149 | Sequence 149, App |
| C 272 | 14.8 | 74.0 | 677 | 13 | US-10-027-632-14827 | Sequence 14827, A |
| C 273 | 14.8 | 74.0 | 677 | 16 | US-10-027-632-14827 | Sequence 14827, A |
| C 274 | 14.8 | 74.0 | 686 | 15 | US-10-029-386-20409 | Sequence 20409, A |
| C 275 | 14.8 | 74.0 | 693 | 13 | US-10-027-632-126342 | Sequence 126342, A |
| C 276 | 14.8 | 74.0 | 693 | 13 | US-10-027-632-127006 | Sequence 127006, A |
| C 277 | 14.8 | 74.0 | 693 | 16 | US-10-027-632-126342 | Sequence 126342, A |
| C 278 | 14.8 | 74.0 | 699 | 9 | US-09-925-301-391 | Sequence 391, App |
| C 279 | 14.8 | 74.0 | 715 | 12 | US-09-978-360A-21 | Sequence 21, App |
| C 280 | 14.8 | 74.0 | 715 | 13 | US-10-315-664-44 | Sequence 44, App |
| C 281 | 14.8 | 74.0 | 715 | 13 | US-10-156-761-5474 | Sequence 5474, App |
| C 282 | 14.8 | 74.0 | 735 | 15 | US-10-027-632-157597 | Sequence 157597, A |
| C 283 | 14.8 | 74.0 | 760 | 16 | US-10-027-632-157597 | Sequence 157597, A |
| C 284 | 14.8 | 74.0 | 760 | 16 | US-10-027-632-126343 | Sequence 126343, A |
| C 285 | 14.8 | 74.0 | 761 | 13 | US-10-027-632-126343 | Sequence 126343, A |
| C 286 | 14.8 | 74.0 | 762 | 16 | US-10-369-493-43164 | Sequence 43164, A |
| C 287 | 14.8 | 74.0 | 789 | 13 | US-10-027-632-127007 | Sequence 127007, A |
| C 288 | 14.8 | 74.0 | 789 | 13 | US-10-027-632-127007 | Sequence 127007, A |
| C 289 | 14.8 | 74.0 | 812 | 13 | US-10-027-632-170926 | Sequence 170926, A |
| C 290 | 14.8 | 74.0 | 812 | 16 | US-10-027-632-170926 | Sequence 170926, A |
| C 291 | 14.8 | 74.0 | 812 | 16 | US-10-027-632-168805 | Sequence 168805, A |
| C 292 | 14.8 | 74.0 | 813 | 13 | US-10-027-632-168805 | Sequence 168805, A |
| C 293 | 14.8 | 74.0 | 813 | 16 | US-10-027-632-168805 | Sequence 168805, A |
| C 294 | 14.8 | 74.0 | 816 | 13 | US-10-027-632-163233 | Sequence 163233, A |
| C 295 | 14.8 | 74.0 | 816 | 13 | US-10-027-632-163233 | Sequence 163233, A |
| C 296 | 14.8 | 74.0 | 816 | 16 | US-10-027-632-163233 | Sequence 163233, A |
| C 297 | 14.8 | 74.0 | 816 | 16 | US-10-027-632-163233 | Sequence 163233, A |
| C 298 | 14.8 | 74.0 | 840 | 13 | US-10-282-122A-13972 | Sequence 13972, A |
| C 299 | 14.8 | 74.0 | 875 | 13 | US-10-027-632-10849 | Sequence 10849, A |
| C 300 | 14.8 | 74.0 | 875 | 16 | US-10-027-632-10849 | Sequence 10849, A |

ALIGNMENTS

RESULT 3
US-10-027-632-9289
; Sequence 9289, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:

US-10-001-857-91/c
; Sequence 91, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 91
; LENGTH: 993
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-91

Query Match 84.0%; Score 16.8; DB 14; Length 993;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCCAAA 20
DB 485 CGCGGCCCGCCCAAA 466

RESULT 2

US-10-001-873-6/c
; Sequence 6, Application US/10001873
; Publication No. US20020160388A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0275
; CURRENT APPLICATION NUMBER: US/10/001,873
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,055
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,496
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-873-6

Query Match 84.0%; Score 16.8; DB 14; Length 1002;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCCAAA 20
DB 559 CGCGGCCCGCCCAAA 540

APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9289
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9289

Query Match 84.0%; Score 16.8; DB 13; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCGAGCCCAACAA 20
Db 170 CGATGGCTCCAGCCCAACAA 189

RESULT 4

US-10-027-632-9290
; Sequence 9290, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9290
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9290

Query Match 84.0%; Score 16.8; DB 13; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCGAGCCCAACAA 20
Db 170 CGATGGCTCCAGCCCAACAA 189

RESULT 5

US-10-027-632-9291
; Sequence 9291, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9291
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9291

Query Match 84.0%; Score 16.8; DB 13; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCGAGCCCAACAA 20
Db 170 CGATGGCTCCAGCCCAACAA 189

RESULT 6

US-10-027-632-9289
; Sequence 9289, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9289
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9289

Query Match      84.0%; Score 16.8; DB 16; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CGAGGCGCCCGCCCAAA 20
    |||||
Db 170 CGATGCGTCCAGCCCAAA 189

RESULT 7
US-10-027-632-9290
; Sequence 9290, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; POLYMORPHISMS IN THE HUMAN GENOME
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9290
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9290

Query Match      84.0%; Score 16.8; DB 16; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CGAGGCGCCCGCCCAAA 20
    |||||
Db 170 CGATGCGTCCAGCCCAAA 189

RESULT 8
US-10-027-632-9291
; Sequence 9291, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; POLYMORPHISMS IN THE HUMAN GENOME
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9289
; LENGTH: 1094
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-9289

Query Match      84.0%; Score 16.8; DB 16; Length 1094;
Best Local Similarity 90.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CGAGGCGCCCGCCCAAA 20
    |||||
Db 170 CGATGCGTCCAGCCCAAA 189

RESULT 9
US-10-282-122A-40410/c
; Sequence 40410, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40410
```

```

; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Treponema pallidum
US-10-282-122A-40410

Query Match      84.0%; Score 16.8; DB 13; Length 1830;
Best Local Similarity 90.0%; Pzed. No. 1.9e-02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGGGCCCGCCACCA 20
Db      660 CGAGGGCCCAATCCCA 541

RESULT 10
US-10-017-161-1677
; Sequence 1677, Application US/10017161
; Publication No. US2003014368A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1677
; LENGTH: 9851
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(9851)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(581)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2244)..(2394)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2475)..(3642)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4153)..(4238)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4840)..(4953)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5064)..(5179)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5444)..(5833)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6503)..(7011)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7503)..(7674)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7925)..(8081)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8179)..(8325)
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (8393)..(8945)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9525)..(9651)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1066)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1068)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1094)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1097)..(1098)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1100)..(1101)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1126)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1129)..(1133)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1135)..(1139)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1141)..(1143)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1189)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1191)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1196)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1213)..(1214)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1245)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1250)..(1251)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1260)..(1269)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1271)..(1274)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
```

NAME/KEY: modified_base
LOCATION: (1278)..(1377)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1438)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1441)
OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1677

Query Match 84.0%; Score 16.8; DB 15; Length 9851;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGCGCCCGCCGCCCCACAA 20
Db 8208 CCAGCGCCCGCCGCCCCACAA 8227

RESULT 11

US-10-292-798-1337
Sequence 1337, Application US/10292798
Publication No. US20030235833A1
GENERAL INFORMATION:
APPLICANT: ASAI, KIYOSHI
APPLICANT: SUWA, MAKINO
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1337
LENGTH: 9851
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: source
FEATURE:
LOCATION: (1)..(9851)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(581)
FEATURE:
NAME/KEY: CDS
LOCATION: (2244)..(2394)
FEATURE:
NAME/KEY: CDS
LOCATION: (2475)..(3642)
FEATURE:
NAME/KEY: CDS
LOCATION: (4153)..(4238)
FEATURE:
NAME/KEY: CDS
LOCATION: (4840)..(4953)
FEATURE:
NAME/KEY: CDS
LOCATION: (5064)..(5179)
FEATURE:
NAME/KEY: CDS
LOCATION: (5444)..(5833)
FEATURE:
NAME/KEY: CDS
LOCATION: (6503)..(7011)

FEATURE:
NAME/KEY: CDS
LOCATION: (7503)..(7674)
FEATURE:
NAME/KEY: CDS
LOCATION: (7925)..(8081)
FEATURE:
NAME/KEY: CDS
LOCATION: (8179)..(8325)
FEATURE:
NAME/KEY: CDS
LOCATION: (8393)..(8945)
FEATURE:
NAME/KEY: CDS
LOCATION: (9525)..(9651)
FEATURE:
NAME/KEY: modified_base
LOCATION: (1066)..(1066)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1068)..(1068)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1094)..(1094)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1097)..(1098)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1100)..(1101)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1126)..(1126)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1129)..(1133)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1135)..(1139)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1141)..(1143)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1189)..(1189)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1191)..(1191)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1196)..(1196)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1213)..(1214)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1245)..(1245)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base

LOCATION: (1250)..(1251)
 FEATURE: a, t, c, g, unknown or other
 NAME/KEY: modified base
 LOCATION: (1260)..(1269)
 OTHER INFORMATION: a, t, c, g, unknown or other
 FEATURE: a, t, c, g, unknown or other
 NAME/KEY: modified base
 LOCATION: (1271)..(1274)
 OTHER INFORMATION: a, t, c, g, unknown or other
 FEATURE: a, t, c, g, unknown or other
 NAME/KEY: modified base
 LOCATION: (1278)..(1377)
 OTHER INFORMATION: a, t, c, g, unknown or other
 FEATURE: a, t, c, g, unknown or other
 NAME/KEY: modified base
 LOCATION: (1438)..(1438)
 OTHER INFORMATION: a, t, c, g, unknown or other
 FEATURE: a, t, c, g, unknown or other
 NAME/KEY: modified base
 LOCATION: (1441)..(1441)
 OTHER INFORMATION: a, t, c, g, unknown or other
 US-10-292-798-1337

Query Match 84.0%; Score 16.8; DB 16; Length 9851;
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCACAA 20
 DB 8208 CCAGCGCCAGCCACAA 8227

RESULT 12
 US-10-006-430-11/c
 Sequence 11, Application US/10006430
 Publication No. US20030113914A1
 GENERAL INFORMATION:
 APPLICANT: Mark J. Graham
 TITLE OF INVENTION: ANTISENSE MODULATION OF CD81 EXPRESSION
 FILE REFERENCE: RTS-0341
 CURRENT APPLICATION NUMBER: US/10/006,430
 CURRENT FILING DATE: 2001-12-10
 NUMBER OF SEQ ID NOS: 90
 SEQ ID NO 11
 LENGTH: 21501
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: exon
 LOCATION: (518)...(812)
 OTHER INFORMATION: exon 1
 NAME/KEY: exon:intron junction
 LOCATION: (812)...(813)
 OTHER INFORMATION: exon 1:intron 1
 NAME/KEY: intron
 LOCATION: (813)...(13608)
 OTHER INFORMATION: intron 1
 NAME/KEY: intron:exon junction
 LOCATION: (13608)...(13609)
 OTHER INFORMATION: intron 1:exon 2
 NAME/KEY: exon
 LOCATION: (13609)...(13723)
 OTHER INFORMATION: exon 2
 NAME/KEY: exon:intron junction
 LOCATION: (13723)...(13724)
 OTHER INFORMATION: exon 2:intron 2
 NAME/KEY: intron
 LOCATION: (13724)...(17291)
 OTHER INFORMATION: intron 2
 NAME/KEY: intron:exon junction
 LOCATION: (17291)...(17292)

OTHER INFORMATION: intron 2:exon 3
 NAME/KEY: exon
 LOCATION: (17292)...(17389)
 OTHER INFORMATION: exon 3
 NAME/KEY: exon:intron junction
 LOCATION: (17389)...(17390)
 OTHER INFORMATION: exon 3:intron 3
 NAME/KEY: intron
 LOCATION: (17390)...(18169)
 OTHER INFORMATION: intron 3
 NAME/KEY: intron:exon junction
 LOCATION: (18169)...(18169)
 OTHER INFORMATION: intron 3:exon 4
 NAME/KEY: exon
 LOCATION: (18169)...(18243)
 OTHER INFORMATION: exon 4
 NAME/KEY: exon:intron junction
 LOCATION: (18243)...(18244)
 OTHER INFORMATION: exon 4:intron 4
 NAME/KEY: intron
 LOCATION: (18244)...(18612)
 OTHER INFORMATION: intron 4
 NAME/KEY: intron:exon junction
 LOCATION: (18612)...(18613)
 OTHER INFORMATION: intron 4:exon 5
 NAME/KEY: exon
 LOCATION: (18613)...(18717)
 OTHER INFORMATION: exon 5
 NAME/KEY: exon:intron junction
 LOCATION: (18717)...(18718)
 OTHER INFORMATION: exon 5:intron 5
 NAME/KEY: intron
 LOCATION: (18718)...(19065)
 OTHER INFORMATION: intron 5
 NAME/KEY: intron:exon junction
 LOCATION: (19065)...(19066)
 OTHER INFORMATION: intron 5:exon 6
 NAME/KEY: exon
 LOCATION: (19066)...(19167)
 OTHER INFORMATION: exon 6
 NAME/KEY: exon:intron junction
 LOCATION: (19167)...(19168)
 OTHER INFORMATION: exon 6:intron 6
 NAME/KEY: intron
 LOCATION: (19168)...(19824)
 OTHER INFORMATION: intron 6
 NAME/KEY: intron:exon junction
 LOCATION: (19824)...(19825)
 OTHER INFORMATION: intron 6:exon 7
 NAME/KEY: exon
 LOCATION: (19825)...(19911)
 OTHER INFORMATION: exon 7
 NAME/KEY: exon:intron junction
 LOCATION: (19911)...(19912)
 OTHER INFORMATION: exon 7:intron 7
 NAME/KEY: intron
 LOCATION: (19912)...(20000)
 OTHER INFORMATION: intron 7
 NAME/KEY: intron:exon junction
 LOCATION: (20000)...(20001)
 OTHER INFORMATION: intron 7:exon 8
 NAME/KEY: exon
 LOCATION: (20001)...(20610)
 OTHER INFORMATION: exon 8
 US-10-006-430-11

Query Match 84.0%; Score 16.8; DB 15; Length 21501;
 Best Local Similarity 90.0%; Pred. No. 1.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGAGGCGCCAGCCACAA 20
 Db 7509 CAAGCGCCAGCCACAA 7490


```
RESULT 13
US-10-091-625-11
; Sequence 11, Application US/10091625
; Publication No. US20030170636A1
; GENERAL INFORMATION:
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF JAGGED 2 EXPRESSION
; FILE REFERENCE: RTS-0244
; CURRENT APPLICATION NUMBER: US/10/091,625
; CURRENT FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 11
; LENGTH: 28000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-10-091-625-11
Query Match      84.0%; Score 16.8; DB 15; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 14
US-10-096-399A-11
; Sequence 11, Application US/10096399A
; Publication No. US20030185829A1
; GENERAL INFORMATION:
; APPLICANT: Koller, Erich
; APPLICANT: Shepherd, Peter J.
; TITLE OF INVENTION: JAGGED 2 INHIBITORS FOR INDUCING APOPTOSIS
; FILE REFERENCE: ISPH-0660
; CURRENT APPLICATION NUMBER: US/10/096,399A
; CURRENT FILING DATE: 2002-03-12
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 28000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-096-399A-11
Query Match      84.0%; Score 16.8; DB 15; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 15
US-10-461-668-11
; Sequence 11, Application US/10461668
; Publication No. US20030207839A1
; GENERAL INFORMATION:
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF JAGGED 2 EXPRESSION
; FILE REFERENCE: RTS-0244
; CURRENT APPLICATION NUMBER: US/10/461,668
; CURRENT FILING DATE: 2003-05-13
; PRIOR APPLICATION NUMBER: US/10/091,625
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 11
; LENGTH: 28000
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
US-10-461-668-11
Query Match      84.0%; Score 16.8; DB 16; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 16
US-10-388-263-381
; Sequence 381, Application US/10388263
; Publication No. US20030228597A1
; GENERAL INFORMATION:
; APPLICANT: Cowsett, Lex M.
; APPLICANT: Baker, Brenda F.
; APPLICANT: McNeill, John
; APPLICANT: Freier, Susan M.
; APPLICANT: Sasnor, Henri M.
; APPLICANT: Brooks, Douglas G.
; APPLICANT: Ohashi, Cara
; APPLICANT: Wyatt, Jacqueline R.
; APPLICANT: Borchers, Alexander
; APPLICANT: Vickers, Timothy A.
; TITLE OF INVENTION: IDENTIFICATION OF GENETIC TARGETS FOR
; TITLE OF INVENTION: MODULATION BY OLIGONUCLEOTIDES AND
; TITLE OF INVENTION: GENERATION OF OLIGONUCLEOTIDES FOR GENE MODULATION
; FILE REFERENCE: ISIS-4503
; CURRENT APPLICATION NUMBER: US/10/388,263
; CURRENT FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 947
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 381
; LENGTH: 28000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-388-263-381
Query Match      84.0%; Score 16.8; DB 16; Length 28000;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGAGCGCCCGCCAGCCCAAA 20
Db      26057 CCAGCGCCCGCCAGCCCAAA 26076

RESULT 17
US-09-999-121-11/c
; Sequence 11, Application US/09999121
; Publication No. US20030039982A1
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; TITLE OF INVENTION: ISOLATED GENOMIC NUCLEOTIDE FRAGMENTS FROM THE p15 REGION OF
; FILE REFERENCE: JR-10,001-US
; CURRENT APPLICATION NUMBER: US/09/999,121
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/244,705
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 37113
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-999-121-11
Query Match      84.0%; Score 16.8; DB 10; Length 37113;
```

```
Best Local Similarity 90.0%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCACCA 20
Db 17233 CAAGGCGCCCGCCACCA 17214

RESULT 18
US-09-814-353-1789
; Sequence 1789, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1789
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 111
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-1789

Query Match 82.0%; Score 16.4; DB 10; Length 475;
Best Local Similarity 94.4%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCACCA 18
Db 286 CGAGGCGCCCGCCACCA 303

RESULT 19
US-09-814-353-8137
; Sequence 8137, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1789
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 111
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-1789

Query Match 82.0%; Score 16.4; DB 10; Length 475;
Best Local Similarity 94.4%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCACCA 18
Db 286 CGAGGCGCCCGCCACCA 303

RESULT 19
US-09-814-353-8137
; Sequence 8137, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
```

```
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8137
LENGTH: 475
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 111
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-8137

Query Match 82.0%; Score 16.4; DB 10; Length 475;
Best Local Similarity 94.4%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCACCA 18
Db 286 CGAGGCGCCCGCCACCA 303

RESULT 20
US-09-814-353-14521
; Sequence 14521, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14521
LENGTH: 565
TYPE: DNA
ORGANISM: Homo sapiens
US-09-814-353-14521

Query Match 82.0%; Score 16.4; DB 10; Length 565;
Best Local Similarity 94.4%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCACCA 18
Db 288 CGAGGCGCCCGCCACCA 305

RESULT 21
```

US-10-275-595A-47
; Sequence 47, Application US/10275595A
; Publication No. US20040078804A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Valda
; APPLICANT: LAL, Preeti
; APPLICANT: YAO, Monique G.
; APPLICANT: BANDMAN, Olga
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: KEARNEY, Liam
; APPLICANT: POLICKY, Jennifer L.
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0772 USN
; CURRENT APPLICATION NUMBER: US/10/275,595A
; CURRENT FILING DATE: 2003-08-13
; PRIOR APPLICATION NUMBER: US 60/201,960
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/202,729
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 60/209,705
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 60/210,149
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: US 60/213,215
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PERL Program
; SEQ ID NO 47
; LENGTH: 3287
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; FEATURE:
; OTHER INFORMATION: Incyte ID No: 1295420CB1
US-10-275-595A-47

Query Match 82.0%; Score 16.4; DB 17; Length 3287;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCAC 18
Db 220 CGAGGCGCCCGCCAC 237

RESULT 22
US-10-108-260A-308
; Sequence 308, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 308
; LENGTH: 3370
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-308

Query Match 82.0%; Score 16.4; DB 16; Length 3370;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCAC 18
Db 312 CGAGGCGCCCGCCAC 329

RESULT 23
US-09-814-353-20078
; Sequence 20078, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20078
; LENGTH: 3592
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4, 3590, 3591, 3592
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20078

Query Match 82.0%; Score 16.4; DB 10; Length 3592;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGGCGCCCGCCAC 18
Db 292 CGAGGCGCCCGCCAC 309

RESULT 24
US-10-085-117-226/c
; Sequence 226, Application US/10085117
; Publication No. US2003023334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226
; LENGTH: 246144
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```
; NAME/KEY: variation
; LOCATION: (1)....(246144)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-226

Query Match      82.0%; Score 16.4; DB 16; Length 246144;
Best Local Similarity 94.4%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGGCGCCCGAGCCGCCACA 19
Db 14025 GAGGCGCTAGCCGCCACA 14008

RESULT 25
US-10-087-192-754/c
; Sequence 754, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: MORRIS, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER.
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 754
; LENGTH: 276276
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-754

Query Match      82.0%; Score 16.4; DB 13; Length 276276;
Best Local Similarity 94.4%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGAGCGCCCGAGCCGCCAC 18
Db 133204 CGAGCGCCCGAGCCGCCAC 133187

RESULT 26
US-10-211-462-112/c
; Sequence 112, Application US/10211462
; Publication No. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynné, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 112

; NAME/KEY: variation
; LOCATION: (1)....(246144)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-226

Query Match      80.0%; Score 16; DB 13; Length 3845;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GCGCCCGAGCCGCCCAA 20
Db 693 GCGCCCGAGCCGCCCAA 678

RESULT 27
US-10-007-926A-110/c
; Sequence 110, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 110
; LENGTH: 3845
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: tyrosine kinase with immunoglobulin and
; OTHER INFORMATION: epidermal growth factor homology domains (TIE)
; OTHER INFORMATION: Gene.
US-10-007-926A-110

Query Match      80.0%; Score 16; DB 15; Length 3845;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GCGCCCGAGCCGCCCAA 20
Db 693 GCGCCCGAGCCGCCCAA 678

RESULT 28
US-10-101-510-4/c
; Sequence 4, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117, 0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3845
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-4
```

Query Match 80.0%; Score 16; DB 15; Length 3845;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20
|||
Db 693 GCGCCAGCCCCACAA 678

RESULT 29

US-10-021-660-34/c
; Sequence 34, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT APPLICATION NUMBER: US/10/021.660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 3845
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-660-34

Query Match 80.0%; Score 15; DB 15; Length 3845;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20
|||
Db 693 GCGCCAGCCCCACAA 678

RESULT 30

US-10-116-802-162/c
; Sequence 162, Application US/10116802
; Publication No. US20030065157A1
; GENERAL INFORMATION:
; APPLICANT: Amy Lasek
; TITLE OF INVENTION: GENES EXPRESSED IN LUNG CANCER
; FILE REFERENCE: PA-0045 US
; CURRENT APPLICATION NUMBER: US/10/116.802
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 60/281,593
; PRIOR FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 519
; SOFTWARE: PERL Program
; SEQ ID NO 162
; LENGTH: 3914
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 3331519CB1
US-10-116-802-162

Query Match 80.0%; Score 16; DB 13; Length 3914;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GCGCCAGCCCCACAA 20
|||
Db 754 GCGCCAGCCCCACAA 739

Search completed: June 20, 2004, 17:44:06
Job time : 127.122 secs

This page Blank (uspto)


```
;
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 64.4%; Score 17.4; DB 3; Length 4403765;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27
DB 4043594 CCGCCGCGCCGAGGATCGTCACCTTA 4043620

RESULT 4
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 64.4%; Score 17.4; DB 3; Length 4411529;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27
DB 4051526 CCGCCGCGCCGAGGATCGTCACCTTA 4051552

RESULT 5
US-08-484-993B-40
; Sequence 40, Application US/08484993B
; Patent No. 5837497
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,993B
; FILING DATE: 09-NOV-1993
; CLASSIFICATION: 424
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
US-08-484-993B-40

Query Match 63.0%; Score 17; DB 2; Length 1701;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25
DB 1318 CACAGCGCTTCAGCATCTTCACCT 1342

RESULT 6
US-08-484-158B-40
; Sequence 40, Application US/08484158B
; Patent No. 5976545
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Pharmaceutical Compositions for
; IMMUNOCONTRACEPTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,158B
; FILING DATE: 07-JUNE-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,223
; FILING DATE: 09-NOV-93
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-93
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-92
; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 32794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
; US-08-484-158B-40
;
Query Match 63.0%; Score 17; DB 2; Length 1701;
Best Local Similarity 80.0%; Pred No: 43;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCOCGGCTTTAGGATCGACACCT 25
      ||| ||| ||| ||| ||| ||| ||| |||
Db 1318 CACCAGCGCTTCAGCATCTTCACCT 1342

RESULT 7
US-08-484-596A-40
; Sequence 40, Application US/08484596A
; Patent No. 5981228
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T., Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,596A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,223
; FILING DATE: 11-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
US-08-484-596A-40

Query Match 63.0%; Score 17; DB 2; Length 1701;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCT 25
Db 1318 CACCGCGCTTCAGCATCTTCACCT 1342

RESULT 8
US-08-480-150A-40
; Sequence 40, Application US/08480150A
; Patent No. 5989550
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocontraception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,150A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/149,223
; FILING DATE: 09-NOV-1993
; APPLICATION NUMBER: 08/012,990
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/973,341
; FILING DATE: 09-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 31745
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6653
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
US-08-480-150A-40

Query Match 63.0%; Score 17; DB 2; Length 1701;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

us-10-624-714-12.rni

Mon Jun 21 09:02:06 2004

APPLICANT: Hsu, Kuang T.
 APPLICANT: Podolski, Joseph S.
 TITLE OF INVENTION: Materials and Methods for Immunocontraception
 NUMBER OF SEQUENCES: 59
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun.
 STREET: 6300 Sears Tower, 233 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/149,223A
 FILING DATE: 09-NOV-1993
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/012,990
 FILING DATE: 29-JAN-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/973,341
 FILING DATE: 09-NOV-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Clough, David W.
 REGISTRATION NUMBER: 36,107
 REFERENCE/DOCKET NUMBER: 31745
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312/474-6653
 TELEFAX: 312/474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1701 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1698
 US-08-149-223A-40

APPLICANT: Hsu, Kuang T.
 APPLICANT: Podolski, Joseph S.
 TITLE OF INVENTION: Materials and Methods for Immunocontraception
 NUMBER OF SEQUENCES: 59
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 6300 Sears Tower, 233 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/458,731
 FILING DATE: 09-NOV-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/012,990
 FILING DATE: 29-JAN-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/973,341
 FILING DATE: 09-NOV-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Clough, David W.
 REGISTRATION NUMBER: 36,107
 REFERENCE/DOCKET NUMBER: 31745
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312/474-6653
 TELEFAX: 312/474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1701 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1698
 US-08-458-731-40

RESULT 9

US-08-458-731-40
 ; Sequence 40, Application US/08458731
 ; Patent No. 6001599
 ; GENERAL INFORMATION:
 ; APPLICANT: Harris Ph.D., Jeffrey D.
 ; APPLICANT: Hsu, Kuang T.
 ; TITLE OF INVENTION: Materials and Methods for Immunocontraception
 ; NUMBER OF SEQUENCES: 59
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/458,731
 ; FILING DATE: 09-NOV-1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/012,990
 ; FILING DATE: 29-JAN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/973,341
 ; FILING DATE: 09-NOV-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clough, David W.
 ; REGISTRATION NUMBER: 36,107
 ; REFERENCE/DOCKET NUMBER: 31745
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6653
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1701 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1698
 ; US-08-458-731-40

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

RESULT 10

US-08-149-223A-40
 ; Sequence 40, Application US/08149223A
 ; Patent No. 6027727
 ; GENERAL INFORMATION:
 ; APPLICANT: Harris Ph.D., Jeffrey D.

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

RESULT 11

US-09-252-991A-1416/c
 ; Sequence 1416, Application US/09252991A
 ; Patent No. 8551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 1416
 ; LENGTH: 2262

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

Query Match 63.0%; Score 17; DB 3; Length 1701;

Best Local Similarity 80.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 CACCGCGCTTTAGGATCGACACCT 25

Db 1318 CACGAGCGCTTCAGCATCTTCACCT 1342

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1416

Query Match      61.5%; Score 16.6; DB 4; Length 2262;
Best Local Similarity 82.6%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 CCGCGCTTTAGATCGACACTG 26
      ||| |||| |||| |||| |||| ||||
DB      2144 CCGAGCTTCAGGTTGCGACGCTG 2122

RESULT 12
US-09-252-991A-1192
; Sequence 1192, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1192
; LENGTH: 2391
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1192

Query Match      61.5%; Score 16.6; DB 4; Length 2391;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      4 CCGCGCTTTAGATCGACACTG 26
      ||| |||| |||| |||| |||| ||||
DB      662 CCGAGCTTCAGGTTGCGACGCTG 684

RESULT 13
US-08-014-153D-37
; Sequence 37, Application US/08014153D
; GENERAL INFORMATION:
; APPLICANT: Hadlock, Kenneth G.
; Goh, Chin-Joo
; Fong, Steven K.H.
; TITLE OF INVENTION: Method and Assay for HTLV
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/014.153D
; FILING DATE: 05-Feb-1993
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/653,091
; FILING DATE: 08-FEB-1991
; APPLICATION NUMBER: US 07/366,313

```

```

; FILING DATE: 13-JUN-1989
; APPLICATION NUMBER: US 06/948,270
; FILING DATE: 31-DEC-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, Allan A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: G4C1P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 369-9500
; TELEFAX: (415) 368-0709
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 3A3B(1), Fig. 8
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-08-014-153D-37

Query Match      60.7%; Score 16.4; DB 1; Length 102;
Best Local Similarity 76.9%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 ACCCGCGCTTAGGATCGACACTGA 27
      |||| |||| |||| |||| |||| ||||
DB      60 ACCCGCGCTTAGGATCGACACTGA 85

RESULT 14
US-07-743-518-21
; Sequence 21, Application US/07743518
; Patent No. 5397696
; GENERAL INFORMATION:
; APPLICANT: YAMAGIHARA, RICHARD
; APPLICANT: NERURKAR, VIVEK R.
; APPLICANT: JENKINS, CAROL
; APPLICANT: MILLER, MARK
; APPLICANT: GARRUTO, RALPH M.
; TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC
; TITLE OF INVENTION: VIRUS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743.518
; FILING DATE: 19910812
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714527 CUSH
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 base pairs
; TYPE: NUCLEIC ACID

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 32..205
US-07-743-518-21

Query Match 60.7%; Score 16.4; DB 1; Length 235;
Best Local Similarity 76.9%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
DB 157 ACCCGCGCTTTAGGATCGAGTCTCTGA 182

RESULT 15

US-08-014-153D-5
Sequence 5, Application US/08014153D
GENERAL INFORMATION:
APPLICANT: Hadlock, Kenneth G.
Filing Date: 05-Feb-1993
TITLE OF INVENTION: Method and Assay for HTLV
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genelabs Technologies, Inc.
STREET: 505 Penobscot Drive
CITY: Redwood City
STATE: CA
COUNTRY: USA
ZIP: 94063

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/014, 153D
FILING DATE: 05-Feb-1993
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/653,051
FILING DATE: 08-FEB-1991
APPLICATION NUMBER: US 07/366,313
FILING DATE: 13-JUN-1989
APPLICATION NUMBER: US 06/948,270
FILING DATE: 31-DEC-1986
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, Allan A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: G4C1P3
TELEPHONE: (415) 369-9500
TELEFAX: (415) 368-0709
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 427 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: p21E(1)CS, Fig. 2A
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Query Match 60.7%; Score 16.4; DB 1; Length 427;
Best Local Similarity 76.9%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
DB 354 ACCCGCGCTTTAGGATCGAGTCTCTGA 379

RESULT 16

US-07-743-518-25
Sequence 25, Application US/07743518
Patent No. 5397696
GENERAL INFORMATION:
APPLICANT: YANAGIHARA, RICHARD
APPLICANT: NERURKAR, VIVEK R.
APPLICANT: JENKINS, CAROL
APPLICANT: MILLER, MARK
APPLICANT: GARRUTO, RALPH M.
TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC
VIRUS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/743,518
FILING DATE: 19910812
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26,581
REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-743-518-25

Query Match 60.7%; Score 16.4; DB 1; Length 520;
Best Local Similarity 76.9%; Pred. No. 74;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
DB 382 ACCCGCGCTTTAGGATCGAGTCTCTGA 407

RESULT 17

US-09-107-532A-1993
Sequence 1993, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham

STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1993:
SEQUENCE CHARACTERISTICS:
LENGTH: 723 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...723
SEQUENCE DESCRIPTION: SEQ ID NO: 1993:
US-09-107-532A-1993

Query Match 60.7%; Score 16.4; DB 4; Length 723;
Best Local Similarity 94.4%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 TTAGGATCGACACTGA 27
Db 594 TTAGGATCGACACTGA 611

RESULT 18
US-09-543-681A-613
Sequence 613, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 613
LENGTH: 1317
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-613

Query Match 60.7%; Score 16.4; DB 4; Length 1317;
Best Local Similarity 76.9%; Pred. No. 82;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACTG 26
Db 1158 CATCCCGCATTAGAAATCGTCACATG 1183

RESULT 19
PCT-US96-00547-54
Sequence 54, Application PC/TUS9600547
GENERAL INFORMATION:
APPLICANT: Virogenetics Corporation
TITLE OF INVENTION: RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS
TITLE OF INVENTION: AND USES
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00547
FILING DATE: 12-JAN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/372,664
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2621
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 1814 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US96-00547-54

Query Match 60.7%; Score 16.4; DB 5; Length 1814;
Best Local Similarity 76.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACTGA 27
Db 1455 ACCCGCGCTTGAAATCGAGTCTGTA 1480

RESULT 20
PCT-US96-00547-41
Sequence 41, Application PC/TUS9600547
GENERAL INFORMATION:
APPLICANT: Virogenetics Corporation
TITLE OF INVENTION: RECOMBINANT POXVIRUS-HTLV, COMPOSITIONS
TITLE OF INVENTION: AND USES
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00547
FILING DATE: 12-JAN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/372,664
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2621
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US96-00547-41

Query Match 60.7%; Score 16.4; DB 5; Length 1892;
Best Local Similarity 76.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACCTGA 27
Db 1508 ACCCGCGCTTTAGGATCGACCTGA 1533

RESULT 21

US-08-286-870A-7
Sequence 7, Application US/08286870A
Patent No. 6083605
GENERAL INFORMATION:
APPLICANT: ELY, S
APPLICANT: TAILOR, RH
APPLICANT: TIPPETT, JM
APPLICANT: BLENK, RG
TITLE OF INVENTION: BACTERIAL GENES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN
ADDRESSER: Intellectual Property Group of
ADDRESSER: PILLSBURY, MADISON & SUTRO LLP
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,870A
FILING DATE: 05-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/520228
FILING DATE: 09-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8910624.9
FILING DATE: 09-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: PAUL N. KOKULIS

REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 70608/220720
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2159 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2159
US-08-286-870A-7

Query Match 60.7%; Score 16.4; DB 3; Length 2159;
Best Local Similarity 76.9%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACCTG 26
Db 916 CACCGCGCTTTAGGATCGACCTG 941

RESULT 22

US-09-121-321-1
Sequence 1, Application US/09121321
Patent No. 6090783
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,321
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/933,803
FILING DATE: 19-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12CIP
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9045 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:

ORGANISM: human
FEATURE:
NAME/KEY: LTR
LOCATION: 1..757
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 8584..8589
FEATURE:
NAME/KEY: LTR
LOCATION: 8278..9032
US-09-121-321-1

Query Match 60.7%; Score 16.4; DB 3; Length 9045;
Best Local Similarity 76.9%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
|||||
Db 6427 ACCCGCGCTTTAGGATCGAGTCTCTGA 6452

RESULT 23

US-08-933-803A-1
Sequence 1, Application US/08933803A
Patent No. 6218522

GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igarashi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku
TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,803A
FILING DATE: 19-SEP-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12CIP
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9045 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: human
FEATURE:
NAME/KEY: LTR
LOCATION: 1..757
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 8584..8589
FEATURE:

NAME/KEY: LTR
LOCATION: 8278..9032
US-08-933-803A-1

Query Match 60.7%; Score 16.4; DB 3; Length 9045;
Best Local Similarity 76.9%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
|||||
Db 6427 ACCCGCGCTTTAGGATCGAGTCTCTGA 6452

RESULT 24

US-09-539-333D-1/c
Sequence 1, Application US/09539333D
Patent No. 6476208

GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Bihain, Bernard
APPLICANT: Essioux, Laurent
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
FILE REFERENCE: GENSET-047AUS
CURRENT APPLICATION NUMBER: US/09/539,333D
CURRENT FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: US 60/126,903
PRIOR FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: US 60/131,971
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/132,065
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/143,928
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 60/145,915
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: US 60/146,453
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/146,452
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/162,288
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: US 09/416,384
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 1

LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 31..1107
OTHER INFORMATION: 5' regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 1108..1289
OTHER INFORMATION: exon A g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 14877..14920
OTHER INFORMATION: exon B g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 18778..18862
OTHER INFORMATION: exon Bbis g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 25593..25740
OTHER INFORMATION: exon C g35018 gene
FEATURE:

NAME/KEY: exon
LOCATION: 29388..29502
OTHER INFORMATION: exon D g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29967..30282
OTHER INFORMATION: exon E g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 64666..64812
OTHER INFORMATION: exon F g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 65505..65853
OTHER INFORMATION: exon G g35018 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 65854..67854
OTHER INFORMATION: 3'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 94124..94964
OTHER INFORMATION: exon g35017
FEATURE:
NAME/KEY: exon
LOCATION: 201188..201234
OTHER INFORMATION: exon S g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216836..216915
OTHER INFORMATION: exon V g35030 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 213818..215818
OTHER INFORMATION: 3'regulatory region g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215941
OTHER INFORMATION: exon R complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215975
OTHER INFORMATION: exon Rbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..216952
OTHER INFORMATION: exon Qbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..217061
OTHER INFORMATION: exon Q complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 217027..217061
OTHER INFORMATION: exon Q1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 229647..229742
OTHER INFORMATION: exon X complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 230408..230721
OTHER INFORMATION: exon P complement g34872 gene
FEATURE:
NAME/KEY: exon

LOCATION: 231272..231412
OTHER INFORMATION: exon Obis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231787..231880
OTHER INFORMATION: exon O2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231870..231879
OTHER INFORMATION: exon O1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 234174..234321
OTHER INFORMATION: exon O complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 237406..237428
OTHER INFORMATION: exon Nbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239807
OTHER INFORMATION: exon N2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239853
OTHER INFORMATION: exon N complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240569
OTHER INFORMATION: exon M1117 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240596
OTHER INFORMATION: exon M1090 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240617
OTHER INFORMATION: exon M1069 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240644
OTHER INFORMATION: exon M52 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240824
OTHER INFORMATION: exon M862 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240994
OTHER INFORMATION: exon M692 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..241685
OTHER INFORMATION: exon M1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240800..240933
OTHER INFORMATION: exon M51 complement g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 241686..243685
OTHER INFORMATION: 5'regulatory region g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 290652..292652
OTHER INFORMATION: 3'regulatory region g34665 gene
FEATURE:
NAME/KEY: exon
LOCATION: 292653..292841

Query Match 60.7%; Score 16.4; DB 4; Length 319608;
Best Local Similarity 76.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OTHER INFORMATION: 8-293-130 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 206545
OTHER INFORMATION: 8-292-198 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 207313
OTHER INFORMATION: 8-251-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208285
OTHER INFORMATION: 8-289-322 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 208960
OTHER INFORMATION: 8-287-249 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 209123
OTHER INFORMATION: 8-287-86 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 209631
OTHER INFORMATION: 8-285-319 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210361
OTHER INFORMATION: 8-283-278 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210463
OTHER INFORMATION: 8-283-176 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 210486
OTHER INFORMATION: 8-283-153 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210583
OTHER INFORMATION: 8-283-56 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 210879
OTHER INFORMATION: 8-282-345 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 210964
OTHER INFORMATION: 8-282-260 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 210979
OTHER INFORMATION: 8-282-245 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 211050
OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGA
NAME/KEY: allele
LOCATION: 211132
OTHER INFORMATION: 8-282-92 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 211247
OTHER INFORMATION: 8-281-367 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 211315
OTHER INFORMATION: 8-281-299 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 211366
OTHER INFORMATION: 8-281-248 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 212520
OTHER INFORMATION: 8-279-197 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 212821
OTHER INFORMATION: 8-278-289 : polymorphic base C or T
NAME/KEY: allele

Query Match 60.7%; Score 16.4; DB 4; Length 319608;
Best Local Similarity 76.9%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTAGGATCGACCTGA 27
DB 120195 ACCCACTTGGAGTCTACTCTCTGA 120170

RESULT 26

US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Query Match 60.7%; Score 16.4; DB 4; Length 1830121;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 GCTTTAGGATCGACCT 25
DB 545007 GCTTTAGGATCGACCT 545024
RESULT 27
US-09-543-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA

; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 888
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-888

Query Match 59.3%; Score 16; DB 4; Length 651;
Best Local Similarity 79.2%; Pred.No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24
Db 97 CAGCGTCGTTTTATGATCGACACC 120

Search completed: June 20, 2004, 11:45:55
Job time : 48.9821 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 733.143 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-12

Perfect score: 27
Sequence: 1 caccgcgcttagatcgacactga 27

Scoring table:
IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.em.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.ov.*
21: em.or.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.un.*
28: em.vi.*
29: em.htg.hum.*
30: em.htg.inv.*
31: em.htg.other.*
32: em.htg.mus.*
33: em.htg.pln.*
34: em.htg.pln.*
35: em.htg.rat.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|--------------|---------------------|
| C 1 | 27 | 100.0 | 17131 | 1 | AE007080 | AE007080 Mycobacte |
| C 2 | 27 | 100.0 | 306550 | 1 | BX248342 | BX248342 Mycobacte |
| C 3 | 27 | 100.0 | 348247 | 15 | BX842579 | BX842579 Mycobacte |
| C 4 | 19.4 | 71.9 | 139421 | 8 | AC087545 | AC087545 Oryza sat |
| C 5 | 19.4 | 71.9 | 300829 | 8 | AE017097 | AE017097 Oryza sat |
| C 6 | 18.6 | 70.4 | 151076 | 5 | BX005194 | BX005194 zebrafish |
| C 7 | 18.6 | 68.9 | 48 | 8 | ATH52463 | ATH52463 Arabidops |
| C 8 | 18.6 | 68.9 | 2000 | 6 | AX509902 | AX509902 Sequence |
| C 9 | 18.6 | 68.9 | 2851 | 1 | AF548445 | AF548445 Unculture |
| C 10 | 18.6 | 68.9 | 11481 | 1 | AE014432 | AE014432 Brucella |
| C 11 | 18.6 | 68.9 | 13916 | 1 | AE009507 | AE009507 Brucella |
| C 12 | 18.6 | 68.9 | 72598 | 8 | AB007545 | AB007545 Arabidops |
| C 13 | 18.6 | 68.9 | 11019 | 10 | AL807391 | AL807391 Mouse DNA |
| C 14 | 18.6 | 68.9 | 162646 | 9 | AC078809 | AC078809 Homo sapi |
| C 15 | 18 | 66.7 | 1287 | 6 | AX122212 | AX122212 Sequence |
| C 16 | 18 | 66.7 | 1287 | 6 | BD164329 | BD164329 Nvel pol |
| C 17 | 18 | 66.7 | 1335 | 6 | AX067039 | AX067039 Sequence |
| C 18 | 18 | 66.7 | 1410 | 6 | AX067037 | AX067037 Sequence |
| C 19 | 18 | 66.7 | 1417 | 6 | AX813887 | AX813887 Sequence |
| C 20 | 18 | 66.7 | 11198 | 1 | AE009079 | AE009079 Agrobacte |
| C 21 | 18 | 66.7 | 14354 | 1 | AE008044 | AE008044 Agrobacte |
| C 22 | 18 | 66.7 | 84615 | 9 | AL354747 | AL354747 Human DNA |
| C 23 | 18 | 66.7 | 86257 | 9 | AL954650 | AL954650 Human DNA |
| C 24 | 18 | 66.7 | 109889 | 10 | AL954830 | AL954830 Mouse DNA |
| C 25 | 18 | 66.7 | 110000 | 2 | LMFCHRI16_02 | Continuation (3 of |
| C 26 | 18 | 66.7 | 110000 | 2 | LMFCHRI16_03 | Continuation (4 of |
| C 27 | 18 | 66.7 | 142340 | 10 | AC124465 | AC124465 Mus muscu |
| C 28 | 18 | 66.7 | 184424 | 2 | AC106319 | AC106319 Rattus no |
| C 29 | 18 | 66.7 | 206112 | 2 | AC084412 | AC084412 Mus muscu |
| C 30 | 18 | 66.7 | 213462 | 2 | AC099282 | AC099282 Rattus no |
| C 31 | 18 | 66.7 | 213474 | 2 | AC103199 | AC103199 Rattus no |
| C 32 | 18 | 66.7 | 223040 | 2 | AC128102 | AC128102 Rattus no |
| C 33 | 18 | 66.7 | 231274 | 2 | AC099393 | AC099393 Rattus no |
| C 34 | 18 | 66.7 | 231509 | 2 | AC131426 | AC131426 Rattus no |
| C 35 | 18 | 66.7 | 232415 | 2 | AC095640 | AC095640 Rattus no |
| C 36 | 18 | 66.7 | 237378 | 2 | AC087038 | AC087038 Mus muscu |
| C 37 | 18 | 66.7 | 246108 | 2 | AC103535 | AC103535 Rattus no |
| C 38 | 18 | 66.7 | 255746 | 2 | AC122622 | AC122622 Rattus no |
| C 39 | 18 | 66.7 | 323450 | 1 | SM5591790 | SM5591790 Sinorhizo |
| C 40 | 18 | 66.7 | 337200 | 1 | AP005280 | AP005280 Corynebac |
| C 41 | 18 | 66.7 | 338579 | 1 | AP003004 | AP003004 Mesorhizo |
| C 42 | 18 | 66.7 | 349980 | 6 | AX127149 | AX127149 Sequence |
| C 43 | 17.8 | 65.9 | 2319 | 1 | BSAVAG | X74289 B.stearothe |
| C 44 | 17.8 | 65.9 | 4839 | 1 | BSORF4 | Y08751 B.stearothe |
| C 45 | 17.8 | 65.9 | 119882 | 8 | NCSE6 | AL670004 Neurospor |
| C 46 | 17.8 | 65.9 | 301550 | 1 | AP005079 | AP005079 Vibrio pa |
| C 47 | 17.6 | 65.2 | 894 | 1 | AF146711 | AF146711 Chromobac |
| C 48 | 17.6 | 65.2 | 1207 | 1 | CHRPHERYD | M55915 Chromobacte |
| C 49 | 17.6 | 65.2 | 2407 | 8 | AK069904 | AK069904 Oryza sat |
| C 50 | 17.6 | 65.2 | 4382 | 9 | HSN802264 | AL137724 Homo sapi |
| C 51 | 17.6 | 65.2 | 4919 | 6 | AX405811 | AX405811 Sequence |
| C 52 | 17.6 | 65.2 | 5135 | 1 | TTNUSAINF | Z49001 T.thermophi |
| C 53 | 17.6 | 65.2 | 121796 | 2 | AC147335 | AC147335 Pan trogl |
| C 54 | 17.6 | 65.2 | 154737 | 9 | AC078834 | AC078834 Homo sapi |
| C 55 | 17.6 | 65.2 | 171452 | 2 | AC122297 | AC122297 Mus muscu |
| C 56 | 17.6 | 65.2 | 190220 | 5 | AL935305 | AL935305 zebrafish |
| C 57 | 17.6 | 65.2 | 241213 | 2 | AC139952 | AC139952 Rattus no |
| C 58 | 17.6 | 65.2 | 249406 | 2 | AC118963 | AC118963 Rattus no |
| C 59 | 17.6 | 65.2 | 256017 | 2 | AC126312 | AC126312 Rattus no |
| C 60 | 17.6 | 65.2 | 298600 | 1 | AP005959 | AP005959 Bradyrhiz |
| C 61 | 17.6 | 65.2 | 302998 | 1 | AE016921 | AE016921 Chromobac |
| C 62 | 17.6 | 65.2 | 306214 | 2 | AC123205 | AC123205 Rattus no |
| C 63 | 17.6 | 65.2 | 325483 | 1 | AP005050 | AP005050 Streptomy |
| C 64 | 17.6 | 65.2 | 340900 | 1 | SM5591791 | AL591791 Sinorhizo |
| C 65 | 17.4 | 64.4 | 538 | 1 | MTU85630 | U85630 Mycobacteri |

Pred. No. is the number of results predicted by chance to have a

| | | | | | | | | | | | | | | | |
|-------|------|------|--------|----|-----------|-----------|-------------|-----------|-----|------|------|--------|---|------------|-----------|
| C 212 | 17 | 63.0 | 186566 | 2 | BX5111265 | Danio rer | BX5111265 | Danio rer | 285 | 16.8 | 62.2 | 260894 | 2 | AC119873 | Mus muscu |
| C 213 | 17 | 63.0 | 186566 | 2 | BX5111265 | Danio rer | BX5111265 | Danio rer | 286 | 16.8 | 62.2 | 261151 | 2 | AC096326 | AC096326 |
| C 214 | 17 | 63.0 | 193317 | 9 | AC098645 | Papio anu | AC098645 | Papio anu | 287 | 16.8 | 62.2 | 279252 | 2 | AC093941 | Rattus no |
| C 215 | 17 | 63.0 | 196238 | 2 | AC132286 | Mus muscu | AC132286 | Mus muscu | 288 | 16.8 | 62.2 | 310850 | 1 | AP005142 | Streptoco |
| C 216 | 17 | 63.0 | 197581 | 2 | AC137043 | Rattus no | AC137043 | Rattus no | 289 | 16.6 | 61.5 | 546 | 6 | AX461531 | Sequence |
| C 217 | 17 | 63.0 | 203050 | 1 | AL646071 | Ralstonia | AL646071 | Ralstonia | 290 | 16.6 | 61.5 | 546 | 6 | AX505992 | Sequence |
| C 218 | 17 | 63.0 | 203830 | 2 | AC096894 | Rattus no | AC096894 | Rattus no | 291 | 16.6 | 61.5 | 546 | 8 | AY081728 | Arabidops |
| C 219 | 17 | 63.0 | 207823 | 10 | AL929142 | Mouse DNA | AL929142 | Mouse DNA | 292 | 16.6 | 61.5 | 546 | 8 | AY125540 | Arabidops |
| C 220 | 17 | 63.0 | 212779 | 9 | AC104031 | Homo sapi | AC104031 | Homo sapi | 293 | 16.6 | 61.5 | 581 | 8 | LLOC010974 | Lilium lo |
| C 221 | 17 | 63.0 | 213119 | 10 | AC091694 | Mus muscu | AC091694 | Mus muscu | 294 | 16.6 | 61.5 | 587 | 6 | BD073325 | Novel nuc |
| C 222 | 17 | 63.0 | 213505 | 2 | AC132571 | Mus muscu | AC132571 | Mus muscu | 295 | 16.6 | 61.5 | 587 | 8 | AF090826 | Lilium lo |
| C 223 | 17 | 63.0 | 213970 | 2 | AC097765 | Rattus no | AC097765 | Rattus no | 296 | 16.6 | 61.5 | 588 | 8 | AF090826 | Lilium lo |
| C 224 | 17 | 63.0 | 214425 | 2 | AC136109 | Rattus no | AC136109 | Rattus no | 297 | 16.6 | 61.5 | 601 | 8 | AB003781 | Lilium lo |
| C 225 | 17 | 63.0 | 215987 | 2 | AC134464 | Mus muscu | AC134464 | Mus muscu | 298 | 16.6 | 61.5 | 1423 | 8 | AF385706 | Arabidops |
| C 226 | 17 | 63.0 | 217768 | 10 | AC074336 | Mus muscu | AC074336 | Mus muscu | 299 | 16.6 | 61.5 | 1444 | 8 | AY060561 | Arabidops |
| C 227 | 17 | 63.0 | 218567 | 2 | AC107735 | Mus muscu | AC107735 | Mus muscu | 300 | 16.6 | 61.5 | 1587 | 8 | BT009450 | Triticum |
| C 228 | 17 | 63.0 | 219107 | 2 | AC141343 | Rattus no | AC141343 | Rattus no | | | | | | | |
| C 229 | 17 | 63.0 | 222528 | 2 | AC107118 | Rattus no | AC107118 | Rattus no | | | | | | | |
| C 230 | 17 | 63.0 | 225079 | 2 | AC126203 | Rattus no | AC126203 | Rattus no | | | | | | | |
| C 231 | 17 | 63.0 | 226832 | 2 | AC110969 | Rattus no | AC110969 | Rattus no | | | | | | | |
| C 232 | 17 | 63.0 | 227709 | 2 | BX571722 | Danio rer | BX571722 | Danio rer | | | | | | | |
| C 233 | 17 | 63.0 | 237343 | 2 | BX569792 | Danio rer | BX569792 | Danio rer | | | | | | | |
| C 234 | 17 | 63.0 | 239377 | 2 | BX511167 | Danio rer | BX511167 | Danio rer | | | | | | | |
| C 235 | 17 | 63.0 | 239377 | 2 | BX511167 | Danio rer | BX511167 | Danio rer | | | | | | | |
| C 236 | 17 | 63.0 | 240886 | 2 | AC096519 | Rattus no | AC096519 | Rattus no | | | | | | | |
| C 237 | 17 | 63.0 | 245358 | 2 | AC136008 | Mus muscu | AC136008 | Mus muscu | | | | | | | |
| C 238 | 17 | 63.0 | 247012 | 2 | AC109008 | Rattus no | AC109008 | Rattus no | | | | | | | |
| C 239 | 17 | 63.0 | 247381 | 2 | AC134193 | Rattus no | AC134193 | Rattus no | | | | | | | |
| C 240 | 17 | 63.0 | 247381 | 2 | AC106199 | Rattus no | AC106199 | Rattus no | | | | | | | |
| C 241 | 17 | 63.0 | 251970 | 2 | AC094791 | Rattus no | AC094791 | Rattus no | | | | | | | |
| C 242 | 17 | 63.0 | 252602 | 2 | AC091704 | Mus muscu | AC091704 | Mus muscu | | | | | | | |
| C 243 | 17 | 63.0 | 258276 | 2 | AC129659 | Rattus no | AC129659 | Rattus no | | | | | | | |
| C 244 | 17 | 63.0 | 261258 | 2 | AC091365 | Rattus no | AC091365 | Rattus no | | | | | | | |
| C 245 | 17 | 63.0 | 264321 | 10 | AL807762 | Mouse DNA | AL807762 | Mouse DNA | | | | | | | |
| C 246 | 17 | 63.0 | 265456 | 2 | AC120829 | Rattus no | AC120829 | Rattus no | | | | | | | |
| C 247 | 17 | 63.0 | 279012 | 2 | AC126134 | Rattus no | AC126134 | Rattus no | | | | | | | |
| C 248 | 17 | 63.0 | 280277 | 2 | AC125932 | Rattus no | AC125932 | Rattus no | | | | | | | |
| C 249 | 17 | 63.0 | 287086 | 2 | BX530407 | Danio rer | BX530407 | Danio rer | | | | | | | |
| C 250 | 17 | 63.0 | 287086 | 2 | BX530407 | Danio rer | BX530407 | Danio rer | | | | | | | |
| C 251 | 17 | 63.0 | 289671 | 2 | AC097291 | Rattus no | AC097291 | Rattus no | | | | | | | |
| C 252 | 17 | 63.0 | 296131 | 2 | AL928718 | Mus muscu | AL928718 | Mus muscu | | | | | | | |
| C 253 | 17 | 63.0 | 300094 | 2 | AC113682 | Rattus no | AC113682 | Rattus no | | | | | | | |
| C 254 | 17 | 63.0 | 303698 | 3 | AB003434 | Drosophil | AB003434 | Drosophil | | | | | | | |
| C 255 | 17 | 63.0 | 304517 | 1 | AB016789 | Pseudomon | AB016789 | Pseudomon | | | | | | | |
| C 256 | 17 | 63.0 | 308760 | 2 | AC126721 | Rattus no | AC126721 | Rattus no | | | | | | | |
| C 257 | 17 | 63.0 | 310539 | 2 | AC118854 | Rattus no | AC118854 | Rattus no | | | | | | | |
| C 258 | 17 | 63.0 | 310539 | 2 | AC118854 | Rattus no | AC118854 | Rattus no | | | | | | | |
| C 259 | 17 | 63.0 | 316050 | 1 | BX321859 | Nitrosomo | BX321859 | Nitrosomo | | | | | | | |
| C 260 | 17 | 63.0 | 325667 | 2 | AC112413 | Rattus no | AC112413 | Rattus no | | | | | | | |
| C 261 | 17 | 63.0 | 329100 | 1 | SME591787 | Sinorhizo | AL591787 | Sinorhizo | | | | | | | |
| C 262 | 17 | 63.0 | 332551 | 2 | AC094116 | Rattus no | AC094116 | Rattus no | | | | | | | |
| C 263 | 17 | 63.0 | 333195 | 2 | AC126635 | Rattus no | AC126635 | Rattus no | | | | | | | |
| C 264 | 17 | 63.0 | 333878 | 2 | AC111474 | Rattus no | AC111474 | Rattus no | | | | | | | |
| C 265 | 17 | 63.0 | 339391 | 2 | AC117149 | Rattus no | AC117149 | Rattus no | | | | | | | |
| C 266 | 17 | 63.0 | 345345 | 2 | AC126167 | Rattus no | AC126167 | Rattus no | | | | | | | |
| C 267 | 17 | 63.0 | 348942 | 1 | BX572596 | Rhodopseu | BX572596 | Rhodopseu | | | | | | | |
| C 268 | 17 | 63.0 | 349970 | 1 | BX571659 | Wolirella | BX571659 | Wolirella | | | | | | | |
| C 269 | 16.8 | 62.2 | 804 | 1 | AB039546 | Pseudomon | AB039546 | Pseudomon | | | | | | | |
| C 270 | 16.8 | 62.2 | 1146 | 1 | BSAMAA | Y08753 | B.stearothe | Y08753 | | | | | | | |
| C 271 | 16.8 | 62.2 | 3625 | 8 | BDNA28RR | X50411 | B.emersonii | X50411 | | | | | | | |
| C 272 | 16.8 | 62.2 | 51385 | 1 | AB014162 | Streptoco | AB014162 | Streptoco | | | | | | | |
| C 273 | 16.8 | 62.2 | 57876 | 9 | AC007074 | Homo sapi | AC007074 | Homo sapi | | | | | | | |
| C 274 | 16.8 | 62.2 | 62736 | 2 | AC103774 | Mus muscu | AC103774 | Mus muscu | | | | | | | |
| C 275 | 16.8 | 62.2 | 112513 | 2 | AC141825 | Apis mell | AC141825 | Apis mell | | | | | | | |
| C 276 | 16.8 | 62.2 | 146258 | 2 | AC023479 | Homo sapi | AC023479 | Homo sapi | | | | | | | |
| C 277 | 16.8 | 62.2 | 148911 | 2 | AL353142 | Homo sapi | AL353142 | Homo sapi | | | | | | | |
| C 278 | 16.8 | 62.2 | 160012 | 9 | AL136231 | Human DNA | AL136231 | Human DNA | | | | | | | |
| C 279 | 16.8 | 62.2 | 176977 | 2 | AC105784 | Rattus no | AC105784 | Rattus no | | | | | | | |
| C 280 | 16.8 | 62.2 | 206801 | 10 | AC073761 | Mus muscu | AC073761 | Mus muscu | | | | | | | |
| C 281 | 16.8 | 62.2 | 217323 | 2 | AC115724 | Mus muscu | AC115724 | Mus muscu | | | | | | | |
| C 282 | 16.8 | 62.2 | 217788 | 2 | AC124577 | Mus muscu | AC124577 | Mus muscu | | | | | | | |
| C 283 | 16.8 | 62.2 | 237565 | 2 | AC134023 | Rattus no | AC134023 | Rattus no | | | | | | | |
| C 284 | 16.8 | 62.2 | 242966 | 2 | AC111931 | Rattus no | AC111931 | Rattus no | | | | | | | |

ALIGNMENTS

RESULT 1
LOCUS AE007080/c 17131 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 166 of 280 of the complete genome.
ACCESSION AE007080 AE000516
VERSION AE007080.1 GI:13882094
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 17131)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 17131)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1..17131
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
104..1357
/gene="MT2390"
/gene="MT2390"
104..1357
/note="similar to SP:P71884 PID:1524272 GB:AL123456; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="PE family protein"
/protein_id="AAK4682.1"
/db_xref="GI:13882095"
/translation="MTANALARKDRHFTTINVPSSLDNRRSGRTSMOFLSVIPE
QVESAQDLAGIRLSASAAAGTTAVUSAEDVSTAIATFAGVGRCOVLISA
QSAFADFENLLKTCATYRTETPANAQSNVLNAPARSLLEGPSAESVQNSAP

```

TLGGHSTVITAGLAAGRAVATVEQQAANAVALPLPSAGAGLAQVAVGVVTVAGGSA
KLATLQGAAPWLRKSGGEFIVAGSALTGVALLQPAUVGVVQAGGFLTAGTSAAATG
LGLTLGAGVFSQGVNLAASGTAATGLGLLSAGVOLPSPAFLLAVPFLALGSGVSL
AIAVQLVQGVQVHLSLVVNVVAGIAALQTAGAQAQVAGVNNHTMLAARQAGPGLVLOT
AGGFPAQOIGHLTITAGNAATVLTIS"
/sequence complement (1392..2939)
/sequence="WT2391"
/sequence complement (1392..2939)
/sequence="WT2391"
/sequence="similar to PID:1813569; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="nitrite extrusion protein"
/protein_id="AAK46683.1"
/db_xref="GI:13882096"
/translation="NEQHTLQREESPRSPAARSLRRLGGRSHITHWDPEDLGAWAG
NKGIAARRLLMSVTVTHLIGYSVTMLPVLLEIMPQDVVGFSTSDKFLLTGTLATLFGAF
LRMPVALASATFGGRNATFSAIVLLIIPAIGTIVLTHFGLPLWPLYVLCALTLGGG
NFASSNANAFYPRHLKSGAGTGGVNGVPAIOVLGLLATAITVGERKPYLVGCAL
YVVLVIAIVGVSFLPMNVQHRVQVNRRLPIVSATLSTRDTWLLSLILYLTGSGFTG
PSFVQVQLQNFLNCCGSPARILHVELAFVGPFLAAVARIYGGRLADRVGSSRLT
LIVFAMTLAAGLLSLSTLGRHVGQHRGATMGVYFVCFVALFVLSGLNGSVYKMI
PTIFACRSRLDLSAERDWSRIISGVYIGVFAAFGALGGVGINMALRESYLSYSGG
TDAFWIFMWCYAAAVALTKVYDRRTVTDMGLQALVROPASTPABELIGPTQSDRF
SGCSISA"
/sequence complement (3173..3700)
/sequence="WT2392"
/sequence complement (3173..3700)
/sequence="WT2392"
/sequence="identified by Glimmer2; putative"
/codon start=1
/transl table=11
/product="hypothetical protein"
/protein_id="AAK46684.1"
/db_xref="GI:13882097"
/translation="MRRORSAPVILALALLALALLALIVLGLGASCAWKPPTRRSPN
TKKSDGPTADTQRAIAAIVPVFGSKWEITRHTNCRHLHWQIIPITIASOSTFQ
QLLFFDRNIPILGSPRNPKPXITVLPAGDDTITVQYQWIGSDQECCTPGTIGTVRPHI
GSDGLKALGSPHQ"
/sequence complement (3574..4161)
/sequence="WT2393"
/sequence="WT2393"
/sequence="identified by Glimmer2; putative"
/codon start=1
/transl table=11
/product="hypothetical protein"
/protein_id="AAK46685.1"
/db_xref="GI:13882098"
/translation="MTVVVLSAARRNRRAPOSVTVTVTVTVPSGRRYADARH
ELSENWFLPFINASGVDPKRVFPVLPQIGRLTPDAVEAGIADDDLPMAARWI
GSRPCLIQPNTMGDNGYLGPLAGQVCDRLVMGASRSTCSRLPVIASVDERLSV
LKPVRPRLHSISFIFKGRPGEVLTITGYNFRGVP"
/sequence complement (4586..6232)
/sequence="WT2394"
/sequence="WT2394"
/sequence="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="malate oxidoreductase"
/protein_id="AAK46686.1"
/db_xref="GI:13882099"
/translation="MSDARVPRIPAALSAPSINRGVGTTHQORRLGLTGRLPSAVLT
LDQAEVYHMQSLATIGRNLLEQLHYRHEVLYFKVLADHLPMLPVPVYPTVGE
AIQRFSDYRGQSLFSLIDSEDETEEAFTNLGLEDVLDIVCTDAEILIGDWG
GGIOIAGVAGLALYTAGGVDPRRCLAVSLDVTQNEQLLADPFYGLNRRHARRRGREYD
EFVSYIETAOBLFPRAILHPEDFGPANAKILDTYGDYCVFENDMQGTGAVLAAV
YSGLKVTGILPRLDQITIVFAGTAGMGIDQIRAMVADGATLECAVSOIWPIDRPGI
LFDMDMDURDPOVPAKRNHQGVAVGVRGLSDAIKASFTILLGSLVYGATKEV
VEAMFASCKHPMIFLSNPTSRMEIAPADVLAWSNRALLATGSPVAPVEFDEITYVI

```

```

GOANNVLAPPGIGLGVIVAGARLITRRMLHAAKAIAHOANPTNPGDSILLPDVQNLRA
ISTVAEAVYRAAVQDGVASRTHDDVRQAIVDTWMLPAYD"
/sequence complement (6186..7799)
/sequence="WT2395"
/sequence complement (6186..7799)
/sequence="WT2395"
/sequence="similar to GP:5881853; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="drug transporter"
/protein_id="AAK46687.1"
/db_xref="GI:13882100"
/translation="MNRQTLLTIATGLGLFIFLDLIVNVALPDIQSFVAGBBDGL
QNVASYSIGMAVFTMSAATLADLYGRRRWYLGVSFTLGLSIACGLAPSIAVLTTAR
GAQGLGAARVSTSLALYSAPPEAKERAKAIGIWTATASIGTTGPTLGLGLVDWQ
WKSIFVNLPMGALVLFILCLVESCHERRARFRLSQLLFIVAGLVAVYVIGPQ
IGWTSVQITVLMNTAAVGCALFEVLERRSNPMDLTLFRDTSYALATICTVFFAV
YCMLLTTOFLQNVRYTPTSVTGLMILPFSAAVAVSPVGLHVGRIGARVPIGLCL
YMLGLLMIIFSEHRSALVVLGLGCSGVCLPTITVMTAVPAERAGASGIM
SAQRAIGSTIGFVGLSVALWLSATLPLEHLERAVDPVQRHVLAEIITIDANPRAHV
GLVPRRHTHERDPVAIAEEDFIEGIRVALLVATATLAVFLAGRWTFPRDVHTAGSD
LSERUPTAMTECAVSHMPGATWCLWPA"
/sequence complement (7826..7996)
/sequence="WT2396"
/sequence="WT2396"
/sequence="identified by Glimmer2; putative"
/codon start=1
/transl table=11
/product="hypothetical protein"
/protein_id="AAK46688.1"
/db_xref="GI:13882101"
/translation="MAICDQTLGCASDGSVRSFRGFPENGRRLRNPQNWGRFTAGVS
QLGEPIPLAHGL"
/sequence complement (8274..9206)
/sequence="WT2397"
/sequence="WT2397"
/sequence="similar to PID:2342631; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="cysteine synthase"
/protein_id="AAK46689.1"
/db_xref="GI:13882102"
/translation="MSIAEDITQLIGRTPLVRLRVTGAVADIVAKLEFFNPANSVK
DRIGVAMLOARQAGLIPKDTITILEPTSGNTGIALMVCARGYCVLTMPETMSLER
RMLRAYGAELITPGADMGSAIAKABELAKTDQRYFYQOFENPAIHRVTAE
EYWRDIDGVLDIIVWAGVGTGTTGVAOVIKERKPSARFVAPEAASPVLSGGQKGP
PIGGIGAGVPPVLDQDLVDELIITVGNEDALAVARLAREEGLLVGISSGAATVAALQ
VARRPENAKLIVVLPDPGERYLSTPLPADVAD"
/sequence complement (9210..9899)
/sequence="WT2398"
/sequence="WT2398"
/sequence="similar to GP:2342632; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="serine acetyltransferase CysE, putative"
/protein_id="AAK46690.1"
/db_xref="GI:13882103"
/translation="MLTAMRGDIRAARDPAATALEVIFCYPGVHAWGHRLLAHWL
WQRELLARAABAEFTRLTGVDIHPGAVIGARVFDHATGVVIGETAEVGGDDVTIYH
CVTLGSSMVGSKRHPTVGDRIIVIGAKVLGPIKIGDSRIGANAVVVKVPPSAVV
VGVGHGVCQSQSPGPPDWRPLDVLVGLSLSLITRVARLDALSGGQAAGVIRPPE
AGIWHGDEFSI"
/sequence complement (10315..11283)
/sequence="WT2399"
/sequence="WT2399"
/sequence="identified by Glimmer2; putative"

```

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46691.1"
```

Query Match 100.0%; Score 27; DB 1; Length 17131;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTGA 27
|||||
Db 3777 CACCGCGCTTTAGGATCGACACCTGA 3751
|||||

RESULT 2
BX248342/c
LOCUS
DEFINITION
BX248342 Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 9/14
ACCESSION
BX248342 BX248333
VERSION
BX248342.1 GI:31619031
KEYWORDS
complete genome.
SOURCE
Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1
AUTHORS
Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrall,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.

TITLE
The complete genome sequence of Mycobacterium bovis
JOURNAL
Online Publication
REMARK
PNAS 10.1073/pnas.1130426100 (Microbiology)
REFERENCE
2 (bases 1 to 306550)

AUTHORS
Garnier,T.
TITLE
Direct Submission
JOURNAL
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw Addlesone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PF4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1. 306550

FEATURES
source
1. 306550
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
201..1286
/gene="adh22"
/locus_tag="Mb2283"
201..1286
/gene="adh22"
/locus_tag="Mb2283"
201..1286
/gene="adh22"
/locus_tag="Mb2283"
/EC_number="1.2.1.1-"
note="Mb2283, adhE2, len: 361 aa. Equivalent to Rv2259,
len: 361 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 361 aa overlap). Probable adhE2,
zinc-containing alcohol dehydrogenase, similar to several,
especially mycothiol-dependent formaldehyde dehydrogenase
from Amycolatopsis methanolica P80094 (360 aa). Contains
P800059 Zinc-containing alcohol dehydrogenases signature.
FASTA scores: >sp|P80094|FADH_AMEYE
NAD|MYCOTHIOL-DEPENDENT FORMALDEHYDE DEHYDROGENASE
(MD-PALDH) Length = 360, Expect = e-156, Identities =
268/358 (74%). Also similar to Rv0162c, (MTC128.02c, 35.0%

gene

CDS

identity in 371 aa overlap)."
/codon_start=1
/transl_table=11
/product="Probable zinc-dependent alcohol dehydrogenase
AdhE2"
/protein_id="CAD97144.1"
/db_xref="GI:31619032"
/translation="MSQTVRGVIARQKGEPELVNI VVDPGPEAAVVDVTACGVCHT
DLTVREGGINDEYFPLGLGHEAAGLIEAVGPGTAVEPGDVLINLRWAVCCQCKRKG
RPYVCFDTEAEOKVTITDGTETALGIGAFADKTLVHSGCTKVDPAADPAVAGLL
CGVMAGLGAATNGVTRDITVAVICGGVGDAIAGAALVGAKRITIAVDITDITKLD
WARTFGATHTNAREVDVQAI GGLTDFGADVIDAVGRPETYQQAFYADLAGTVV
LVGPTDMDLMDPLVDFFSHGGLKSSWYSGDCLPESDFTLLDLYQGRPLQRFVYS
ERIGLEDVEAFHMGKGKVLRSVWML"
1286..1921
/locus_tag="Mb2284"
1286..1921
/locus_tag="Mb2284"
/note="Mb2284, -, len: 211 aa. Equivalent to Rv2260, len:
211 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 211 aa overlap). Conserved
hypothetical protein, similar to hypothetical proteins
Rv0634C, Rv1637C, Rv3677C, Rv2581C from Mycobacterium
tuberculosis and to various hydrolases. FASTA scores:
spt|O6154|O6154 HYPOTHETICAL 21.3 KD PROTEIN (200 aa)
opt: 355, E(): 4e-15; (37.4% identity in 198 aa
overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97145.1"
/db_xref="GI:31619033"
/translation="MAAIEKVTHTGTFDGGSEVDNINLWDDSEVVVFDAAHHA
APIIDVAGRGKRVAVITCTGHNDHVTVAPELFDALDPAVLMHFGDVLMTFPGKSF
RAVSDGDVAVRGVGTETALRTPGHSPGVCWYAPELGPGTGTGTFSGDTLPAGPGATG
RSYSDRPTILRSISGRGLPAGTGVVHTGHGDDTTIGDEIVHYEEWVARGH"
complement (1998..3506)
/locus_tag="Mb2285c"
complement (1998..3506)
/locus_tag="Mb2285c"
/note="Mb2285c, -, len: 502 aa. Equivalent to Rv2262c and
Rv2261c, len: 360 aa and 140 aa, from Mycobacterium
tuberculosis strain H37Rv, (94.7% identity in 357 aa
overlap and 100.0% identity in 140 aa overlap). Conserved
hypothetical protein, with function unknown but some
similarity to N-terminal 70% of
P23930|P77703|LNT_ECOLI|CUTE|B0657 APOLIPOPROTEIN
N-ACYLTRANSFERASE (EC 2.3.1.-) from Escherichia coli
strain K12 (512 aa), FASTA scores: opt: 239, E(): 1.6e-07,
(30.4% identity in 359 aa overlap). Note that neighboring
ORF shows similarity to N-terminal part of PC6803
apolipoprotein N-acyltransferase from Synecocystis sp.,
suggesting possibility of frameshift. Sequence of clones
from two sources has been checked but no error found.
Appear to be two extra bases at position 1876970 compared
to CDG1551 strain. Conserved hypothetical protein, with
function unknown but some similarity to C-terminal end of
PC6803 apolipoprotein N-acyltransferase from
Synecocystis sp. Note that next ORF shows similarity to
N-terminal part of P74055 APOLIPOPROTEIN N-ACYLTRANSFERASE
from Escherichia coli (519 aa), FASTA scores: opt: 142,
E(): 0.007, (29.9% identity in 117 aa overlap), suggesting
possible frameshift. Sequence of clones from two sources
has been checked but no error found.
REMARK-M bovis-M tuberculosis: In Mycobacterium
tuberculosis strain H37Rv, Rv2262c and Rv2261c exist as 2
genes. In Mycobacterium bovis, a 2 bp deletion (ct-*)
results in a single product which is more similar to
Rv2262c."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD97146.1"

gene

CDS

gene

CDS

FT penicillin-binding membrane protein, similar to many
FT bacterial PBP2 protein e.g.
FT P11882|PBP2_NEIME|PENA|NMA2072|NMB0413|penicillin-binding
FT protein 2 (pbb-2) from Neisseria meningitidis (serogroups A
FT and B) (581 aa), FASTA scores: Opt: 665, E(): 1.6e-31,
FT (33.2% identity in 591 aa overlap); etc. Also similar to
FT Rv0016c and Rv2866c from Mycobacterium tuberculosis
FT (2.8e-10). Contains F50017 possible ATP/GTP-binding site
FT motif A (P-loop) near C-terminus. FASTA best: PBP2_NEIME
FT P11882|penicillin-binding protein 2 (pbb-2). (581 aa) opt:
FT 665, E(): 1.6e-31; (33.2% identity in 591 aa overlap)"
FT /transl_table=11
FT /function="Involved in peptidoglycan biosynthesis."
FT /gene="pbbB"
FT /locus_tag="Rv2163c"
FT /product="Probable penicillin-binding membrane protein
FT pbbB"
FT /protein_id="CAB08664.1"
FT /translation="MSRAAPRRASQSTRPARGLRPPGAQVEGVRKPGTKOKARQA
FT QEAQKSRATRSADVAPAGSTRATROVDVGTGASFFVHRHTGNVILVLMVAAT
FT QLFQVSHAGLRQAAGQKLVYDVQPAAGSIVDRNDSLAFTIEARLTFQPKIR
FT RQLEBARKTSRAQDQORLDIAQEAQVAGLNKPNDAALUKLOSDETFVYLARVDP
FT AVASAIKAYPEVGAERDLQRYPGSLAANVGGIDWDGHLGLSDSLAVLAGTDS
FT SVTYDRGSDGVVPGSYNRHKAHVGSTVLTLDNDIQFYVQQVQQAQNLGSAHVSA
FT VLLDAKTEVLAMANDTFDSQDIGROGDKGNPAVSPFEGSVNKLVAAGAVTEH
FT GLSSDPELVGQSTQMGVTVHDAWEGHVEYTTTGVGSSNVNLTLSQVGPERY
FT YDMRKFCLGQSTGVLGPGESAGVPPIDQSGSTFANLPIGQSLTLMQTMGYQAI
FT ANDGVVPRIKAIKVAADGRTSEPRDDIRVYSAQTAQIVRQMLRAVORDPMGQIQ
FT GTGTVAGVFGQMAKTAQOINPGCGCFYDVIWITFAGIATADNPRYVIGIMLNDP
FT ARNSDGAAGSAALPFIHNIAGLMQRENVLPSPDGPPLVQAT"
FT complement(470. .493)
FT /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
FT /complement(12245. .3399)
FT /note="Rv2164c. (MTCY270.04), len: 384 aa. Probable pro-
FT rich conserved membrane protein, equivalent to
FT M0907|AL022602 putative conserved membrane protein from
FT Mycobacterium leprae (377 aa) (AL022602), FASTA scores:
FT Opt: 1495, E(): 1.7e-56, (62.217% identity in 397 aa
FT overlap)."
FT /transl_table=11
FT /function="UNKNOWN"
FT /locus_tag="Rv2164c"
FT /product="PROBABLE CONSERVED PROLINE RICH MEMBRANE PROTEIN"
FT /protein_id="CAB08663.1"
FT /translation="WRAKREAPKSRSSRRRRADSPAAARTTTNSAPRRIRSRAGK
FT TSAQROARVSRPGPOTSPMLSPDRPAPAKNTSOAKARAKAKAPKLVPPMERL
FT AARLTSLDRPLANKVPFVVLVIGSLGVLGLTLMSTDAERSYQLSNARETRML
FT QCKREALDRVREASAPALAEARRQMIETRTDLHLVQDPGNVWVYVGPKPADGVP
FT PPLNTKLUPEDPPPPKPAAPVLEPVVRVTPGDDPAPPAKSPSEVIVRTPDGATGG
FT ATHLPQTQGLPQGPVPIPGAPMPAPPLGAPSPAPAPENVPVLOVGAAPAGLPGPA
FT FVAATPGLSGSQMPVAPPAPVPANGQFGVTPAVPTAPGAPR"
FT complement(3396. .4586)
FT /note="Rv2165c. (MTCY270.03), len: 396 aa. Conserved
FT hypothetical protein, shows strong similarity to several
FT hypothetical bacterial proteins but has extra 80 aa
FT residues at N-terminus FASTA best: YLXA_BACSU Q07876
FT hypothetical 35.3 kDa protein in ftsl (311 aa) opt: 781,
FT E(): 0, (45.6% identity in 296 aa overlap), BELONGS TO THE
FT YABC (E. COLI), YLXA (B. SUBTILIS) FAMILY"
FT /transl_table=11
FT /function="UNKNOWN"
FT /locus_tag="Rv2165c"
FT /product="CONSERVED HYPOTHETICAL PROTEIN"
FT /protein_id="CAB08662.1"
FT /translation="WQTRAPWSPLPTEALVAPPNARFVSSRDGLGAGAPGIAASRSTAC
FT /TGGITVADPGSGPTGPHVYVLAORCFELLTPALTRYVDPGSCVALLDATTGAGGHA
FT BRFEGLGLRLIGLDRDPTALDVAARSLVFAEDRLTVHTRVYDCLGAALASRYAVRG
FT SYDGLFELVSSQLDRAERGAFTADPLDMDPTTLTAADIYNTVDEAALADIL
FT RYGEERFARRIKVRRKARFYATSTAEVLVLYQAIIPAPRAPVGGHFAKRTFOALR
FT IAVNDELESLETPAALDAIGRIATVAYQSLDRIVKRVFAEVAAPAGLPEVE
FT LPGHEPRSTHTGAERASVAIEIRNPRSTFVRLRALQRVHRAQSQWATEKGDG"
FT complement(4588. .5019)
FT /note="Rv2166c. (MTCY270.02), len: 143 aa. Conserved

FT hypothetical protein; shows strong similarity to several
FT hypothetical bacterial proteins such as YLLB_BACSU P55343.
FT Is equivalent to Mycobacterium leprae hypothetical protein
FT M0905 (143 aa, 92% identity) MLCB268.11c
FT >sp|O69561|YL66 MYCLE HYPOTHETICAL 16.1 KDA PROTEIN M0905
FT >gi|3080482|emb|CAAL18677.1|(AL022602)
FT >gi|13092975|emb|CAC31286.1|(ALS93920). FASTA scores:
FT M0905|M0905 conserved hypothetical protein (143 aa) opt:
FT 873, E(): 3.1e-52; 92.254% identity in 142 aa overlap;
FT YLLB_BACSU P55343 hypothetical 16.6 kDa protein (143 aa)
FT opt: 340, E(): 3.6e-17; (35.0% identity in 143 aa overlap).
FT BELONGS TO THE YABB (E. COLI), YLLB (B. SUBTILIS), MG221 (B.
FT M. GENTILIUM) FAMILY"
FT /transl_table=11
FT /function="UNKNOWN"
FT /locus_tag="Rv2166c"
FT /product="CONSERVED HYPOTHETICAL PROTEIN"
FT /protein_id="CAB08661.1"
FT /translation="MELGYVTPLDKDGLTLPKAPKDALAGLMVTKSQDHSLVYVPR
FT AAPEQLARASAKAPRNPAPAFRLNLAAGTDEQHPDSCRTLSADHRRYASLSKDCV
FT VIGADYLEIRWDQAWQNTQQIHEENFSAASDEALGDIIF"
FT complement(5278. .5305)
FT /note="28 bp Inverted repeat at the left end of IS6110;
FT GAGTCTCCGACTCACCGGCGGCTTCA"
FT complement(5278. .6632)
FT /note="IS6110-6. len: 1355 bp. Insertion sequence IS6110."
FT /insertion_seq="IS6110-6"
FT complement(5320. .56360)
FT /note="RV2167c. (MTCY270.01), len: 346 aa. Probable IS6110
FT transposase. FASTA best: TRA9 MYCTU P19774 putative
FT transposase for insertion sequence (identical)"
FT /transl_table=11
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT ELEMENT IS6110."
FT /locus_tag="RV2167c"
FT /product="PROBABLE TRANSPOSASE"
FT /protein_id="CAAL17494.1"
FT /translation="ABALAAGORRIAKGRDFKRVGLRGRAPASTLITRPIADHOG
FT HRGPGDLRWGVESICTOLTELGVPATPYVDHINRPSRRRLGELKEHLSRVHAA
FT NYGVGARKVWLTLNREGIEVARCTVERLTKLGISGTRGKARRITLADPATRAPDL
FT VQRFGPPAPNRLVADLIYVWAGVFAIVFDIYARRILGRVAGVGSYDNLDAI
FT EQAIWTRQEGVLDLKDVIHHTDRGSQYTSIRFSELAEGIQPSVAGVGSYDNLDAI
FT TNLGYLTLLIKFGKWPWSIEDVELATARVDFWDFNRRILYQYCGDVPVPELEAAVYQAR
FT QRPAAQ"
FT complement(6255. .6581)
FT /note="RV2168c. (MTV021.01c), len: 108 aa. Probable IS6110
FT transposase. FASTA scores: O08155|O08155 HYPOTHETICAL 12.0
FT kDa PROTEIN (108 aa) opt: 697, E(): 0, (100.0% identity in
FT 108 aa overlap). TBPase score is 0.928."
FT /transl_table=11
FT /function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT ELEMENT IS6110."
FT /locus_tag="RV2168c"
FT /product="PROBABLE TRANSPOSASE"
FT /protein_id="CAAL17472.1"
FT /translation="MSGSSRRYPPELRERAVRMVAIRGOHDSWAALISEVARLLGVG
FT CAETVKWVRQAQVDAGAPGTTTSEAEKELRDLNRLRANILKATASAFFAAEEL
FT RPAR"
FT complement(6605. .6632)
FT /note="28 bp Inverted repeat at the right end of IS6110,
FT TGAACCCGCCGCGATGCCGAGATC"
FT complement(6726. .7130)
FT /evidence=EXPERIMENTAL

Query Match 100.0%; Score 27; DB 15; Length 348247;

Best Local Similarity 100.0%; Pred No. 0.023; Length 348247;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
|||||

Db 179460 CACCCGCGCTTTAGGATCGACACCTGA 179434
|||||

```

RESULT 4
AC087545
LOCUS
DEFINITION
    Oryza sativa (japonica cultivar-group) DNA linear PLN 30-MAY-2002
    nbx0019M20, complete sequence.
ACCESSION
    AC087545
VERSION
    AC087545.3 GI:21263269
SOURCE
    HTG.
ORGANISM
    Oryza sativa (japonica cultivar-group)
    Oryza sativa (japonica cultivar-group)
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
    Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
    AUTHORS
    TITLE
    JOURNAL
    AUTHORS
    Llaica, V., Young, S., Kavchok, S., Clark, T., Charydczac, G., Choi, V.,
    Nevill-Manning, C. and Messing, J.
    Direct Submission
    Submitted (08-JAN-2001) The Plant Genome Initiative at Rutgers -
    Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
    Piscataway, NJ 08873
    Chromosome 10
    3 (bases 1 to 139421)
    Llaica, V., Young, S., Kavchok, S., Ward, K., Charydczac, G. and
    Messing, J.
    Direct Submission
    Submitted (02-MAR-2001) The Plant Genome Initiative at Rutgers -
    Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
    Piscataway, NJ 08873
    Chromosome 10
    4 (bases 1 to 139421)
    Llaica, V., Song, R., Young, S., Kavchok, S., Ward, K. and Messing, J.
    Direct Submission
    Submitted (30-MAY-2002) The Plant Genome Initiative at Rutgers -
    Waksman Institute, Rutgers University, 190 Frelinghuysen Road,
    Piscataway, NJ 08873, USA
    On May 30, 2002 this sequence version replaced gi:13184927.
    This BAC overlaps with rice BAC nbx0073D04 (AC087542) and
    nbx0012D19 (AC087546).
FEATURES
    source
    1..139421
    /organism="Oryza sativa (japonica cultivar-group)"
    /mol_type="genomic DNA"
    /sub_strains="japonica"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosomes="10"
    /clone="nbx0019M20"
ORIGIN
    Query Match 71.9%; Score 19.4; DB 8; Length 139421;
    Best Local Similarity 95.2%; Pred. No. 1.5e+02;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
    QY 6 GCGCTTTAGGATCGACACCTG 26
    Db 18797 GCACCTTTAGGATCGACACCTG 18817
RESULT 5
AE017097
LOCUS
DEFINITION
    Oryza sativa (japonica cultivar-group) DNA linear PLN 06-JUN-2003
    77 of the complete sequence.
ACCESSION
    AE017097
VERSION
    AE017097.1 GI:31432234
SOURCE
    Oryza sativa (japonica cultivar-group)
    Oryza sativa (japonica cultivar-group)

```

REFERENCE
AUTHORS
CONSRM
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES
source

misc_feature

gene

mRNA

CDS

gene

mRNA

CDS

gene

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 300029)

The Rice Chromosome 10 Sequencing Consortium
In-depth view of structure, activity, and evolution of rice
chromosome 10

Science 300, 1566-1569 (2003)

2 (bases 1 to 300029)

Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.

Direct Submission

Submitted (05-MAY-2003) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

This is the pseudomolecule for rice chromosome 10, which was
constructed by resolving discrepancies between overlapping BACs,
trimming the overlap regions, and linking the unique sequences to
form a contiguous sequence. Genes in individual BAC clone were
identified by a combination of several methods: Gene prediction
programs, searches of the complete sequence against a peptide
database and EST databases. Genes with similarity to other proteins
are named after the database hits. Genes without significant
peptide similarity but with EST similarity are named as unknown
proteins. Genes without protein or EST similarity, that are
predicted by more than two gene prediction programs over most of
their length are annotated as hypothetical proteins. Genes
encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/addy/tRNAscan-SE/).

Location/Qualifiers

1..300029

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosomes="10"

<1..47797

/notes="Chromosome Sequence Derivation: nucleotide sequence
in this region was derived from BAC clone nbx0073D04
(GB:AC087542)."

<3344..>4619

/locus_tag="OSJNBA0073D04.33"

/notes="Predicted by fgenesh"

join(<3344..3349,4158.>4619)

/locus_tag="OSJNBA0073D04.33"

join(3344..3349,4158..4619)

/locus_tag="OSJNBA0073D04.33"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAP53890.1"

/db_xref="GI:31432235"

/translation="MEIDDKGKRPMSBELTEKHANAVQIAKNEPPPPGVAKVAVQ
GFRKETGDACAGIIVRDRCGLVLLAACKLPRCSATQAEALACLEGVELATNWHMP
IILESDNADVAGLITQASRAEWGGIIAEIRVAMQCLLQVQHVHKRDSNRG"

<6000..>6671

/locus_tag="OSJNBA0073D04.34"

/notes="Predicted by fgenesh"

<6000..>6671

/locus_tag="OSJNBA0073D04.34"

6000..6671

/locus_tag="OSJNBA0073D04.34"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAP53891.1"

/db_xref="GI:31432236"

/translation="MGASASSGERRRRRCGSAIIVADTVSEHFDLTIDGYSHIKAMP
TGKCATSPPTAGCHEWCVDFYPNGKAAAGADMIQFFLRERAAKAAKAAVAAQ
NRFDLLGDSGPGAGSGTGRVLSFPAVDGQGSWVRDFAEQAGCAXDDFTVRFA
VTVFRGRTAAAPESRRPAPSPASPLPMCSYTCVTDCQTRNRSAPTRPSSSTD
GGCQV"

complement(<7340..>10632)

/locus_tag="OSJNBA0073D04.1"

/notes="similar to cinnamyl alcohol dehydrogenase

GB:AA19536.GI:19849248 (Lolium perenne); EST C72961,

mRNA
C24898, BM421887 from this gene"
complement [join(<7340, .7779,7864, .8531,8806, .8919,
10384, .>10632)]
/locus tag="OSJNBa0073D04.1"
complement [join(7568, .7779,7864, .8531,8806, .8919,
10364, .10632)]
/locus tag="OSJNBa0073D04.1"
/codon_start=1
/product="putative cinnamyl alcohol dehydrogenase"
/translation="MATAEAAAARRGGGSGNRGASGAGRRORRAAAGARTAG
GCEAAAAGGAGGVRGRRGAVGCPGAAAPPPRRTHRRSTGGDDVAIKIL
FCGICHSDHICIKNEWHSIYPLVGEIAGVTEVGKNTFRKAGDRVGVGCMWSC
RSCSNNGFHCPEGVFTVNSVDKGTWYGGYSVMVVFHFRVPEAMELDVG
APLICAGITVPMKYIAGNAPKGVHGLGGLGHVAVFAFAFGKLVKIVSSPGK
KREALERIGAVFVSSAEARSTMDGVINTVSANTPMPAPYLALLKPNKMILV
GLFNPLEVFPFSLVHGNITAGSNIAGMADTQEMIELAAKHGVTADIEVIGADVNT
AMERLAKADVRYFRFVIDVGNTHAAAAE"
/db_xref="GI:31432237"
/locus tag="OSJNBa0073D04.36"
/note="similar to retinitis pigmentosa GTPase regulator
GB:AG00551 GI:9937379 (Homo sapiens)"
join(<14775, .15042,15172, .15370,15464, .15728,
15842, .>15934)
/locus tag="OSJNBa0073D04.36"
join(14775, .15042,15172, .15370,15464, .15728,15842, .15934)
/codon_start=1
/product="putative regulatory protein"
/translation="MRDEPTAIRRRRELDSGESRRRQQAEGEGNDGATRGFLAV
RSTRFELDSVGLDGTPEAGDVLVRSSDGGEHTASDGNRRRSGNSLGLRP
KAKRERAPKPSASSNAPVLPALEGQRDRDRAGROKHGGKEWKKGKRLGEDDA
GEESKGECKGLPICRFGKKEEVSCEAAEGGMLPESLGLRLAEGRWNTTAM
TAGRFAERRRRARAEADGGDRAAGDRVAQTRDRTGGNHAGARANERGER"
complement(<16478, .>23409)
/locus tag="OSJNBa0073D04.35"
/note="similar to gag-pol protein GB:AAL59229 GI:18092337
(Zea mays)"
complement [join(<16478, .17166,17233, .21196,21628, .22437,
23383, .>23409)]
/locus tag="OSJNBa0073D04.35"
complement [join(16478, .17166,17233, .21196,21628, .22437,
23383, .23409)]
/locus tag="OSJNBa0073D04.35"
/codon_start=1
/product="putative gag-pol protein"
/translation="MPTVMGCEMAVNAVSFSSHGFGEHHVAQLRLARFLGFCPEY
TGYSDVRPFHCELSVLYIPRGHIGAGLRPHFIVARPTLQIAYQDLSWTALRLAH
DYSRLGSGAPQCLPRLPSGSDARPCPYSESQPLVRLVWELSEAOATQDILLKAY
RSLARRYAALAEARVETGVSVDTI SWDSGLCHASHLSHSSRGSAITPSGSPSS
RAPYSPVSPVPHVPSYTPGHPVASSRGCFIETARKVYVGSVTFRFLHAPP
PPAMASPRRAACAAPAFGEGBGVKPEANESNEQSHRSNHRITVSNKAEQPC
RSNKSHTASDMEQSHRSHTASHNSGEECLPSPHVHNNQHLMVQTOLOGI
AAAVAGYKNSHNGHGMNNSKLTDFLRSPPEFSQTPVEADMDQLKVDVRSIT
LVQCTPVEKTLIAHQLEGPAAWNYCNAPHEPTNIADFEATAPRAAHVESTID
MKCEKFNRLKQGSVNEYLSMFKLARYALEEVDSDKKIRKFLIAGVAGLQLLA
HDPTFOHINKALLIEDARKEATEYK3KSNHQNSRGAAPRYGQPMOYHOSVT
QANOPGVAPQONAPQOCHAPSGNTAPSVTSFKSPQSPASVQCFRCQNGHY
ARQCFQPTNTHGANGSTARTPTPAAQSRUSTQSGSRASNSHFRGCVNHLIRA
ETKQADPVGVSVNSVPAVLFDGASHFSISOAFVKNRKNKTQNLNRPVHISP
GRNIRAQICEVNLRIEVDFLAKPVLDSQSLDILGMDLWAKHQDIDCAEKIT
LQCGPGQVTRPTPTASRSILTCLOVTSLESPIVCEYDPVPEELRTPMDPRIE
FAIELAPCTAIKRPYQMAANELTEVKROI GELSKGYVRPSSPMGAPVLLVKKD
GSEMWIDYALNEVTIKNKYPLRIDDLFDQLGARVFSKIDLSGSHOLKIRSDI
PKTASRYGLYETVMSFGLTNAPTFMNLKIFMEYLDQVYVFIIDILLIYSKE
EEHAEHLINLEKRDQLQFAKSKCFWLDRAVFLGHVISSNGVEVDPKSVAVLAW
NPPKNVSEIFSLGAGYRRFIIEGSKLAPMTLELLKKEKFRWSAACEDSFOEMKK
RLTTAPVLTPDIRKDFEIFODASROGLGCVLIQERKVAYASRQLRHEVNTPTDL

gene
/locus tag="OSJNBa0073D04.39"
/note="Predicted by genemarker"
join(<25617, .25780,25922, .>26162)
/locus tag="OSJNBa0073D04.39"
join(25617, .25780,25922, .>26162)
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP53895.1"
/db_xref="GI:31432240"
/translation="MSVWEPTCAVKGASHRSVAKAATGRMTTPARKKKGKRGKGA
CFPAGSGKRRRVQRLTAAATGSRSAATKARGLQREARMEWRGSGTRAGSQGAAAR
LKAAGORDRAAQTRRTDAGNHAGARANERGER"
/db_xref="GI:31432240"
/locus tag="OSJNBa0073D04.37"
/note="similar to nuclear speckle type protein SP0P
GB:CAA04199 GI:2695708 (Homo sapiens)"
join(<27867, .28919,29563, .29682,30319, .30366,30460, .30580,
31647, .31703,32683, .32791,32872, .33097,33182, .33298,
33615, .33693,33877, .33899,34272, .34316,34446, .>34604)
/locus tag="OSJNBa0073D04.37"
join(27867, .28919,29563, .29682,30319, .30366,30460, .30580,
31647, .31703,32683, .32791,32872, .33097,33182, .33298,
33615, .33693,33877, .33899,34272, .34316,34446, .34604)
/codon_start=1
/product="putative speckle-type protein"
/protein_id="AAP53896.1"
/db_xref="GI:31432241"
/translation="MSSPSQOPASAIKAPTSTGYHRLRIDYVRSISGPTGALSRRD
FVVGROWFISYVNGNRPENAEFISVFLCLDSSSPKPMLOVITFDDEAKKQSLR
KAPVITAFACAGVHVRKRDLDLARSKRIRDFGFTTICDVLSDHIDFTAQDEPVFV
SVPPERDLGLDITSGGDVVFQVGGEAFATHRGLLAARSPLAALYIGPMMEGG
GLQGVYAIKIDMDPLFKALLRYATISLPQMQQGELEEGRAVAQHLLAADRYG
MERLRLLCBAQLCKHIEVASVASILLIADQHCGLKNAACPFELKSPGFAAAMATQE

Query Match 71.9%; Score 19.4; DB 8; Length 300029;
Best Local Similarity 95.2%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 GCCTTTAGCATGCACACCTG 26
|||
Db 56593 GCACCTTTAGCATGCACACCTG 56613

RESULT 6
BX005194/c
LOCUS
DEFINITION
Zebrafish DNA sequence from clone CH211-188A7 in linkage group 25,
complete sequence.
BX005194
Accession
BX005194.10 GI:34996469
VERSION
KEYWORDS
SOURCE
ORGANISM
Danio rerio (zebrafish)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 151076)
AUTHORS
Giselle H.
Direct Submission
Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,

ERPNLELELYSDRGQOQFRKI FAPGPEHEPYGATCVAGGHOIGFNDLKTIEVRDFVL
 AIAGQKTHADFRGVEVQKTVETIYASSRERQWKKI"
 complement (1332..2108)
 /locus tag="BR1344"
 complement (1332..2108)
 /locus tag="BR1344"
 /note="identified by match to TIGR protein family HMM
 TIGR01187"
 /codon start=1
 /transl_table=1
 /product="iron compound ABC transporter, ATP-binding
 protein"
 /protein_id="AAN30258.1"
 /db_xref="GI:23348186"
 /translation="MTLLSVKNDLVILGKRALADASFETKGGFGLVGNAGKTT
 LIRAIAGLVASSQVSLAGRDRLHMGAAEKARSLAYLPQERDVAFVPMVLSIGRS
 ALKVPAGLDRODAIEIEMDMVDSYFGRSMELSGGERARVLIARVLAQDTPVI
 LADPEVGLDPAHLQTLMTFAELAREGRTVIASLHLSLAQHCSRLILLDHGCMAA
 DGTPEVFTPORLDVYGRARIMVMDGSEFVHPTALQPR"
 complement (2105..3106)
 /locus tag="BR1345"
 complement (2105..3106)
 /locus tag="BR1345"
 /note="identified by match to PFAM protein family HMM
 PF03772"
 /codon start=1
 /transl_table=1
 /product="iron compound ABC transporter, permease protein"
 /protein_id="AAN30259.1"
 /db_xref="GI:23348187"
 /translation="MTERHFLLLAGLALLVILFAVSLTGPAAIGIGSFKALLG
 DQDMTVLWRIIRPALLALLIGASLGALQGLYNLPAPGLLVGVSASALG
 AVTAIYSGLSLPPALPILLAGAFVSVELKLAGRNAGTAVILAGVATVSLAGA
 LITALNLSPNPAMEIYFWMGLSLADSENTHALVIPPILLGMLLSLGRSDLSL
 TUGSDAATGVSGLRQVFAVLGRACVGASTAVAGSIGFGLVPHLLRVLGARP
 SRLVAGGIGGALLDAADILRVVMPGRELKGLVITALLIGAFVLMVLFVTRQLV
 complement (3099..3938)
 /locus tag="BR1346"
 complement (3099..3938)
 /locus tag="BR1346"
 /note="identified by match to PFAM protein family HMM
 PF01497"
 /codon start=1
 /transl_table=1
 /product="iron compound ABC transporter, periplasmic iron
 compound-binding protein"
 /protein_id="AAN30260.1"
 /db_xref="GI:23348188"
 /translation="MKRVSVKVTBRLCLAAGLALFAPLPANAADIQBVSVNVCTDQ
 LALLAREQLOSVSLSDPQSLVWADKAERLPINHAQAEVFLQKDDVLVAGTFSS
 RPTGLRLVIRVEFAPARFEDIQELHRIEGLLQGAQAQOIDMKAAALAIK
 RPDKARVLYANSYSGQGLVDEAVLAGLENAGEIGITGSALLPLEKLVLEKP
 DILVRSRRAPALAPENFQHPALRALEKQARAI SMADNLTVCGGPPFSVEAVELAEAE
 ARD"
 complement (3952..5814)
 /locus tag="BR1347"
 complement (3952..5814)
 /locus tag="BR1347"
 /note="identified by match to PFAM protein family HMM
 PF04052"
 /codon start=1
 /transl_table=1
 /product="iron compound TonB-dependent receptor, putative"
 /protein_id="AAN30261.1"
 /db_xref="GI:23348189"
 /translation="MHILSALLASACATLSPALAQDGKDDGVLTDTIVTPPLR
 RASLQRTSVSVDAADIERSAADPLQLQTVSGVTKNGQGSANIMYRGM
 SKOTVVLVNGVRTASGTALANPLTSIERIARGAHSSOYGAIDAIGGVNIIT
 KOGACGERACWGSVGTGSHPKGVSGLQSRSDGDIDYVGAFTGTQGYDTTP
 EAFCHPDDGFLQGSFNALSKDFWKGKLYADGLFSSGRNOYDAPAFNEADSTFP
 TGVKTRIDHTADWSTVEFSGIDNSRNRFKIGSGDRFETRRYGVFASTKSPDTG
 KVSHVTVNGVAREKINTIDYDGTDLAAVFGQSLSEYDALRFDGIRYDHNGVL
 GNVTTYNLGASVEILPDLYLRSYATGTGFRAPTTLYEPGFANPDLOPEKSRSEVGL

NWQATASLDMALYQTRLSDAIMSTAPSVIPYNTASAKVTGLEATLSHRENEOMGKIK
 GMDLKRVDDESGNDLPYRERFKAAAEVNFVPEKLDLTARVLVGGSYTAKNTKK
 LGDYVTADFVALYSIDKQSLFSEVENIFDKOYETSSGVAFQRTIITGLITRN"
 complement (6103..6258)
 /locus tag="BR1348"
 complement (6103..6258)
 /locus tag="BR1348"
 /note="identified by Glimmer2; putative"
 /codon start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAN30262.1"
 /db_xref="GI:23348190"
 /translation="MSSRSIARFAGAVANPQSYCYRNFBQTVPPLRSMKREHGEAIGLK
 PLRPPQL"
 complement (6308..7420)
 /locus tag="BR1349"
 complement (6308..7420)
 /locus tag="BR1349"
 /note="similar to GP:15158527; identified by sequence
 similarity; putative"
 /codon start=1
 /transl_table=1
 /product="ABC transporter, permease protein, putative"
 /protein_id="AAN30263.1"
 /db_xref="GI:23348191"
 /translation="MRIANFRILGKELRGLMRDPMLLIIVVAFTLAIYTAAKAMPE
 TLANAAIIVDEQDSQVSGRTITAFYPPVFPKLISSYEMDNMDGDLTALNIPP
 DFGDILLAGAPTQIANDATMSOAFGTGGVQSVIVAGESEFINHVSAPSAPVDL
 NLSARNGKNGFGAINTVSISSYTMLSIILTGALIRERBHGTHILLVMPVPL
 IMISKWMSGLVNVASAFALTAVQGLSVINGSLIFLAGAAQLFATTSGLIFL
 ATVAGNPGQGLLMLLPLQVLSGATTPRESMPEIITIMLAENTHFVILAOSIL
 FRAGAGVWVQPFALFVIGSVLFLFALRRFRAFLR"
 complement (7422..10190)
 /locus tag="BR1350"
 complement (7422..10190)
 /locus tag="BR1350"
 /note="similar to GP:15158528; identified by sequence
 similarity; putative"
 /codon start=1
 /transl_table=1
 /product="ABC transporter, ATP binding/permease protein"
 /protein_id="AAN30264.1"
 /db_xref="GI:23348192"
 /translation="MSSVFGKAGDGGFVVRARGVDLSYGATRALRNVLDVAGGVNG
 LIQPDGVGSSLSLIAGARTIQOQHVEVLGGDIADKRRQSAIPRIAYNPQGLKQL
 YPTLSVENIDFFGRFGHDAERERRIADLLARTGSAFANRPAKLSGGMKQKTSL
 CCSLIIDPDVLDIPEPTGVDRPLSRQFWELIIDIRKDRPMSVIVATYMBEABERFD
 WLVAMDGOILATGTQELLERTDTPNLDAAFIALLPERQGRHREIHIPAPASDT
 AIALEAHTVRFSDTAVDNVSFRIIPGEI FGLGSGCGKTTMKVLTGLLPAGEG
 TALFREVDPKIDIRRVAGVMSQAFSLYSELTVQNLEHARLFGMDPDTVAREIK
 ELTARFDLADMDLEPLGIRQRLSLAVALIHSPDILDEPSTGSDVPVARDAFW
 QIILADSRDNVTI FVSTHFMNABECLDRISMHAGKVLISIDTPKAITSRNAATLED
 AFVAFQDAI GEQSAPOKIEKDSAEPAQHRKAWPSLNLNRRFATYRREALLK
 RPIRATLALGSLVILMFGVINGLDVEKLTFAVLDRDOTTISREYTOIAGSYRFI
 EKPITDYDOLDRMREGEISLAIETPPNFAANMGROVEVAAIDGMPTRAETAK
 GVQGVHANVLTOKARALYGDATAGNENLEIRVYNPQVSLVAVVAVIPLILLMI
 PMLVLSVVRXELGSLINFYVTVTKFEFLGKOLPVLQGLAFNLMFLAPVAFIF
 RVPTFSVTVTAALLYIITSMGLVISTFMNSQIAAIFGTALMTLIPAVQYSGLI
 DPVSALOGVGAFIGKIYPTTYFTVISRGTFSKALDFSLALGSLPLILLIAPVLLMLGLI
 VLLKQAS"
 complement (10187..11254)
 /locus tag="BR1351"
 complement (10187..11254)
 /locus tag="BR1351"
 /note="similar to GP:15158529; identified by sequence
 similarity; putative"
 /codon start=1
 /transl_table=1
 /product="Hlyd family secretion protein"
 /protein_id="AAN30265.1"
 /db_xref="GI:23348193"

```
Query Match      58.9%; Score 18.6; DB 1; Length 11481;
Best Local Similarity 84.0%; Pred. No. 4e-02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGCGCCTTTAGGATCGACACCTGA 27
      ||||| ||||| ||||| ||||| |||||
Db 8852 CCGCGCCACGGGATCGACACCTGA 8876

RESULT 11
AE009507/c
LOCUS
DEFINITION
  Brucella melitensis 16M DNA linear BCT 20-MAR-2003
  complete sequence.
ACCESSION
  AE009507 AE008917
VERSION
  AE009507.1 GI:17982577
KEYWORDS
SOURCE
  ORGANISM
    Brucella melitensis 16M
  Brucella melitensis 16M
  Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
  Brucellaceae; Brucella.
  1 (bases 1 to 13916)
REFERENCE
  AUTHORS
    DelVecchio,V.G., Kapral,V., Redkar,R.J., Patra,G., Mujer,C.,
    Los,T., Ivanova,N., Anderson,I., Bhattacharya,A., Lykidis,A.,
    Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,
    Mazur,M., Goltzman,E., Selkov,E., Elzer,P.H., Hagius,S.,
    O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kypides,N. and
    Overbeek,R.
  The genome sequence of the facultative intracellular pathogen
  Brucella melitensis
  Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
  11756688
REFERENCE
  2 (bases 1 to 13916)
  DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
  Direct Submission
  Submitted (13-NOV-2001) Institute of Molecular Biology and
  Medicine, University of Scranton, Scranton, PA 18510, USA
  3 (bases 1 to 13916)
  Elzer,P.H. and Hagius,S.
  Direct Submission
  Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
  Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA
  4 (bases 1 to 13916)
  Kapral,V., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A.,
  Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
  Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Haselkorn,R.,
  Kypides,N. and Overbeek,R.
  Direct Submission
  Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
  Park Drive, IL 60612, USA
  5 (bases 1 to 13916)
  Letesson,J.-J.
  Direct Submission
  Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire,
  Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
  61 rue de Bruxelles, Namur 5000, Belgium
  6 (bases 1 to 13916)
  O'Callaghan,D.
  Direct Submission
  Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
  Kennedy, Nimes 30900, France
  1.13916
  /organism="Brucella melitensis 16M"
  /mol_type="genomic DNA"
  /strain="16M"
  /db_xref="taxon:224914"
  /chromosome="I"
  complement(78..458)
  /gene="BMEI0650"
  complement(78..458)
  /gene="BMEI0650"
  /codon_start=1

FEATURES
  source
    gene
    CDS
      825..2093
      /gene="BMEI0651"
      825..2093
      /gene="BMEI0651"
      /codon_start=1
      /transl_table=11
      /product="Hypothetical Cytosolic Protein"
      /protein_id="AAL51832.1"
      /db_xref="GI:17982578"
      /translation="MMGASIGERITQMQLSGOLRALLGLRYQGLVDFCKEWPYAPR
      SRGCDTGLKEKARHREIFVQPGYKVLVQRHRLSHSSSLTCNENSIRQCWFPMFGN
      GGSGQFGTDLVPDCWKNIRRG"
      825..2093
      /gene="BMEI0651"
      825..2093
      /gene="BMEI0651"
      /codon_start=1
      /transl_table=11
      /product="Hypothetical Cytosolic Protein"
      /protein_id="AAL51832.1"
      /db_xref="GI:17982579"
      /translation="NNLNFLAGMKGNRVFADCFATWATVATVINLDFSKQSTAIAD
      GFKRVIAQKSELLAIRDRAKLGINAVIFOVSPTADAFYQSSILPKSSVLTGLVGDPR
      GFPLFMIAEAKHRIEYHAFNPNVSMDSRETQEAALKNSSPSPSPVYKTHPKW
      IGVAKYRVLDPGIPAVRDWGNVIGEVQKVDVDCIQEDDVEYSTPFSQLDDRTFF
      ARYGTRENSKADWRRYNTKLVQDVNFGVRVKIRPTVRFGLSPGGMNRNRKDDPESDT
      LAGATYNSDFADTRRWKDGIVDYIAPQVYWPLEREAAPIYIVKWAQTVSGTVD
      LYTGMALYKAGETRYRLEPQWAAEGEFTIRRELELNEVSEVPEVKGSLLFRESFLREPKL
      RKITDYLRNMGCAHKAQAKKAPF"
      2166..2402
      /gene="BMEI0652"
      2166..2402
      /gene="BMEI0652"
      /codon_start=1
      /transl_table=11
      /product="Hypothetical Protein"
      /protein_id="AAL51833.1"
      /db_xref="GI:17982580"
      /translation="MLIVSLPVSSTRAGSPGCGORPYSYGTPTAESARELLRRFVRO
      SAGAVSIATISISKHPRLSPPCAGGICGKGYA"
      2711..3706
      /gene="BMEI0653"
      2711..3706
      /gene="BMEI0653"
      /codon_start=1
      /transl_table=11
      /product="periplasmic component of efflux system"
      /protein_id="AAL51834.1"
      /db_xref="GI:17982581"
      /translation="MQAMNGSLPDGIASNGRVEAVIDISTKSPGRIEIPANEGS
      FVKAGDVLARMDTQLESYRQAKAQLRRABIGIDTAQSLVTQROAEHAAATVAQR
      EAQDAQRRLAPRSQLESRTVSQQLDDDRATAGAAVGAQAQLAATDAIGA
      AKQYVDAGASVEAKAAATAEADRLATILAPRGVQYRVAQGVLSAGRVLN
      LVLSVSTWTFITTAQAGRVAIGADRLDAAPQYVPAKVSVADVADVAQTFPKTVE
      SEERQKLMFRVKAKIPQILLQKYIQQVKTLGFGVAYVKLAPDAGMFKNLAETVK"
      3935..5698
      /gene="BMEI0654"
      3935..5698
      /gene="BMEI0654"
      /codon_start=1
      /transl_table=11
      /product="ABC TRANSPORTER ATP-BINDING PROTEIN"
      /protein_id="AAL51835.1"
      /db_xref="GI:17982582"
      /translation="MPTSATSLRALPICRKGLARTFIPRCSRILISSGSLAT
      TENGSCRIADLLARTGSAFANRPAKLGCGMKQKTSLLCCSLIHDPDLLIDEPPTG
      VPLSRQFWELIDDIRKDRPGMSVIVATAYEEAEERFVLMVMDQGLATGPQEL
      LERTDTPNDDAIFALLPBEQRGRHREIHIPRASPADAIAIEAEHVTVRFGDFTAV
      DNVSFRIPRGEIFGFGNGCGKTTMKVLTGLTPASEGTAFLFGREVPKIDIRRH
      VGYMQAFSLYSALTQHNLHARLFMGPEPDVARRIKELTARFLADVMDELDPAM
      PLGIRQLSLAVALTSPDILILDEPTSGVDVARDAFQIADLISRRDNVTIYVSTH
      FMNEALCDRIISHMAGKVLISDTPKAITESNAATLEDAFVAFIQDAIGEGEQAPQ
      KLEKQSAEPKQHRKAWPSWLNLSRMFAYTRREALKRDPIRATLAILGSLVLMFV
      IGYGNLDVEKLTFAVLDKDDTTISREYTOQIAGSRYFIEKPPITDYYDLDRMRREGE
      ISLAIEIPNFANVVRGAAG"
      5598..6473
      /gene="BMEI0655"
      5598..6473
      /gene="BMEI0655"
      /codon_start=1
```

/gene="BMEI0655"
/codon_start=1
/transl_table=11
/product="ABC TRANSPORTER ATP-BINDING PROTEIN"
/protein_id="AAL51836.1"
/db_xref="GI:17982583"
/translation="MTISTGECAKASAPLKYRRIISOLMSGGRQVEVAWIDGAMP
TRAETVGVQVQVHANLWTKARLYGDAATAGNENLIRYRNPVDVOSLVAMVPAVI
PILLAMLVREKEVLSIINFTVPTKFEELGKOLPYICLAFNLMLT
AFATVFRVPTGSPVVTATALLVITTSKGLVISTFMSQIAAIGFTALMTLPA
VOYGLIDVLSALQGVAFIGIKYPTTIVTISCGTFSKALDVSALAGSFLPLIAIP
VLMILGIVLLKQAS"
6475..7587
/gene="BMEI0656"
6475..7587
/gene="BMEI0656"
/codon_start=1
/transl_table=11
/product="DAUNORUBICIN RESISTANCE TRANSMEMBRANE PROTEIN"
/protein_id="AAL51837.1"
/db_xref="GI:17982584"
/translation="WRIANIPLRIKELRGLMRDPMMLILIVYAFTLAIYTAAKAMPE
TANNAIAIVDEQSPVSGRIITAFVPPVFPKLISSVEMDNRMDEGLDTFALNIPP
DFORDLAGRATIQINVDATMSQAFTGGGVQSVIVAGEVSEFLNHYEGSAPAPVDL
NURARFQNLKNGFALNTNVISSVTMLSIILITGAALRERHGTIEHLVWPVPIE
IMISKVMSGLVVLVASAFALTFVVOGLSVPIGSIILFLAGAAQLQFATISGIFL
ATVAGSPQFGLMLLILPLQVLSGATTPRESMEPIITQINLAAPNTHFVILAQSIL
FRGAGLVVWPQFLALFVIGSVLFLALRRFAFLR"
8144..9943
/gene="BMEI0657"
8144..9943
/gene="BMEI0657"
/codon_start=1
/transl_table=11
/product="METAL CHELATE OUTER MEMBRANE RECEPTOR"
/protein_id="AAL51838.1"
/db_xref="GI:17982585"
/translation="MAQDGDGDDGVLTIVTPLRRASSLQRTSSVSVVIDRADIE
RSPAQLQLQYSGISVKTNGQGSADIIRMGSSQITVLLVNGVRTASATSGST
ALANIPLTIERIARGASQYGAADIGVINIITQGGACGRACGVSSTGVSH
PMGYASGLQSGSDIGDYAVGAFTQGGDFTTPEAFGEHPDDGFLQGSFNFAL
SGDFDKGTVKADIFRGRNDYAPAFNEADSTAFKGTGTRIDHTADWSSTVEFS
TGDINSENFRKGIEGSDWETRYGVFASTKSFDTGVKVTGCVGVEYREIKNTTI
DYDETGRDLAIVGQVLSYDALRFDGGIRYDHNGQFNVTTYNLGAIVEILLPDLVLR
SVATGFRATFNELTYPGFNDPLQPEKRSVEVGLNMQATASLDMALYQTELSL
AINSTAPYIPYNIAKVTGLEATLSHFNQWIKGMDLRKRPVDSGNDLYRE
RFXAAEVNFKPEKDLTLVLYGGSRVYNKTKLGDYVTFADVALYSIDKQSQL
KFSVENIFDKDYETSTSGVYAPGRTITIGLTRNF"
9957..10796
/gene="BMEI0658"
9957..10796
/gene="BMEI0658"
/codon_start=1
/transl_table=11
/product="METAL CHELATE PERIPLASMIC BINDING PROTEIN"
/protein_id="AAL51839.1"
/db_xref="GI:17982586"
/translation="MKRVSVKWTRELCIAGLALFAPLPANAADIAQVWSINVCTDQ
LAMLAREGQLQSVYSLRDPQLSVWADKAERLPINHAQAEVFLQKPDVLVAGTFSS
RATVGLLRIGIRVEEFAPARSFDOEHRLRIGELLGQRAEQAIQIDDMKALAAIK
RPHAKRVLYANYSVSGGTLVDVAVLGNWAGIEGITGSGALLPLEKLVLEKP
DILVRSRDRAPALAFENFQHPALRALEKRAISNADNLTVCGGPFVSVEVAELAE
ARD"
10774..11790
/gene="BMEI0659"
10774..11790

Query Match 68.9%; Score 18.6; DB 1; Length 13916;
Best Local Similarity 84.0%; Pred. No. 4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 CCGCGCGTTTGGAGTACACACCTGA 27
|||||
5042 CCGCGCGCACGGATCGACACCTGA 5018
|||||

RESULT 12
AB007645
LOCUS
DEFINITION
AB007645
ACCESSION
AB007645.1
VERSION
AB007645.1
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

27268 bp DNA linear PLN 27-DEC-2000
Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K8K14.
AB007645 BA000015
AB007645.1 GI:2564045
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (sites)
Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 5. III.
Sequence features of the regions of 1,191,918 bp covered by
seventeen physically assigned P1 clones
DNA Res. 4 (6), 401-414 (1997)
98162728
2 (bases 1 to 72698)
Nakamura,Y.
Direct Submission
Submitted (03-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1332-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=K8K14
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
Splice predictor (Volker Brendel, Stanford University,
http://grm1.ni.zoel.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K3G17 and the 3' clone is K919.
Location/Qualifiers
1..72698
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="K8K14"
/clone_lib="Mitsui TAC"
complement(join(283..653,770..1619))
/note="contains similarity to unknown protein
gene_id:K8K14.1
sp|Q10058"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09014.1"
/db_xref="GI:9758428"
/translation="MAVISLPPVLTTRITICNSKSKMAKQIGETSKRVFV
CGGVIGVCTAYFLAKKGIATLVEKSAVCAAGKAGGFLAFDWCDSVVAHLARAS
FLHRSIAELANGVESYGYALTVLSVTESKGGGLGLPWNVNGPAKSTIGTT
QTAAQVHPQLFTRKLLSTAREKYGIVTCKLEEVRESRVNSVLEGGRVIDTV
VLAMPWSKFLSSIFRYVCTKAHSIVLEKPEPAITHALFTRPAHGEALDP
EVPRTGEVYICGMSQEEVDDADQVTSNQIYQVLRVAKTVTSYLSNEALVKA

REFERENCE 6 (bases 1 to 162646)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Mar 1, 2002 this sequence version replaced gi:17977513.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

Location/Qualifiers
 1. .162646
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-600G3"
 1. .2004

misc_feature 1. .2004
 /note="overlaps bases 79511..81510 of clone AC117486"
 /function="clone overlap"

complement(248..339)
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 491..556
 /rpt_family="MIR"

repeat_region 7013..7275
 /rpt_family="AluSx"
 repeat_region 7276..7892
 /rpt_family="L1MC3"
 repeat_region complement(8146..8365)
 /rpt_family="MIR"
 repeat_region 9447..9479
 /rpt_family="AT-rich"
 repeat_region 9648..9780
 /rpt_family="GA-rich"
 repeat_region 10440..10660
 /rpt_family="MER20"
 repeat_region complement(10703..11014)
 /rpt_family="AluSx"
 STS 11431..11683
 /standard_name="79189"
 repeat_region complement(11911..12013)
 /rpt_family="MIR"
 repeat_region 12654..12964
 /rpt_family="AluSx"
 repeat_region 13022..13142
 /rpt_family="MIR"
 repeat_region 15041..15337
 /rpt_family="AluSg"
 repeat_region 15456..15480
 /rpt_family="(TTTA)n"
 repeat_region complement(16012..16151)
 /rpt_family="MIR"
 repeat_region 17138..17449
 /rpt_family="AluSp"
 repeat_region 17765..17895
 /rpt_family="L2"
 repeat_region 18594..18860
 /rpt_family="MIR"
 repeat_region 18907..19141
 /rpt_family="MIR"
 repeat_region 19474..19515
 /rpt_family="(TA)n"
 repeat_region complement(19940..20118)
 /rpt_family="AluSx"
 repeat_region 20726..21393
 /rpt_family="L1M3e"
 repeat_region complement(21402..21704)
 /rpt_family="L1"
 repeat_region 21710..23051
 /rpt_family="L1M2"
 repeat_region 23045..24606
 /rpt_family="L1M1"
 repeat_region 24611..24765

Query Match 68.9%; Score 18.6; DB 9; Length 162646;
 Best Local Similarity 84.0%; Pred. No. 3.9e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACT 25
 Db 21237 CACCCGCGCTTTAGGATCGACACT 21261

RESULT 15

AX122212/c

LOCUS AX122212 1287 bp DNA linear PAT 11-MAY-2001

DEFINITION Sequence 2128 from Patent EP1108790.

ACCESSION AX122212

VERSION AX122212.1 GI:14038928

KEYWORDS

SOURCE Corynebacterium glutamicum

ORGANISM Corynebacterium glutamicum

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE 1

AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.


```

TITLE Novel polynucleotides
JOURNAL Patent; EP 1108790-A 2128 20-JUN-2001;
          KIOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
  source
    1. .1287
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1287;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 90 CAGGCGCTCTTTAGGAGCGACACTG 65

RESULT 16
BD164329/c
LOCUS BD164329 Novel polynucleotide. 1287 bp. DNA linear PAT 17-JAN-2003
DEFINITION Novel polynucleotide.
ACCESSION BD164329
VERSION BD164329.1 GI:27870141
KEYWORDS JP 2002191370-A/2128.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1287)
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoo, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotide
JOURNAL Patent: JP 2002191370-A 2128 09-JUL-2002;
COMMENT KYOWA HAKKO KOGYO CO LTD
OS Corynebacterium glutamicum
EN JP 2002191370-A/2128
PF 09-JUL-2002
PD 15-DEC-2000 JP 2000405096
PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI,
PI KEIKO OCHIAI,
PI HARUHIKO YOKOI, NAOKO TATEISHI, AKIHIRO SENOO, NASATO IKEDA, AKIO
PI OZAKI
PC C12N15/09, C12N15/09, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC
C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P7/40, C12P13/
PC 04, C12P13/08,
PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC
G01N33/566,
PC G01N33/569, G01N33/68, G01N37/00, C12P21/08, (C12N1/21, C12R1:15),
PC (C12N1/21, C12R1:13), (C12N1/21, C12R1:01), (C12P13/08, C12R1:15),
PC C12N15/00,
PC C12N5/00, C12N15/00
CC Novel polynucleotide
FH Key
FT source 1. .1287
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
FEATURES
  source
    1. .1287
      /location/Qualifiers
      /organism="unidentified"
      /mol_type="genomic DNA"
      /db_xref="taxon:32644"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1287;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 90 CAGGCGCTCTTTAGGAGCGACACTG 65

TITLE Novel polynucleotides
JOURNAL Patent; EP 1108790-A 2128 20-JUN-2001;
          KIOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
  source
    1. .1287
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1287;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 90 CAGGCGCTCTTTAGGAGCGACACTG 65

```

```

RESULT 17
AX067039/c
LOCUS AX067039 1335 bp. DNA linear PAT 24-JAN-2001
DEFINITION Sequence 621 from Patent WO0100805.
ACCESSION AX067039
VERSION AX067039.1 GI:12544747
KEYWORDS
SOURCE
  Corynebacterium glutamicum
  Corynebacterium glutamicum
  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
  Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in
  membrane synthesis and membrane transport
JOURNAL Patent: WO 0100805-A 621 04-JAN-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
  source
    1. .1335
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
      101. .1315
      /notes="unnamed protein product; FRXA02233"
      /codon_start=1
      /transl_table=11
      /protein_id="CAC26747.1"
      /db_xref="GI:12544748"
      /db_xref="RENTREMBL:CAC26747"
      /translation="MLVSTTWGTVHGDGKIEPGAVAPKRLSGWRTIGIGMORHVI
      AMFGATLLVPTLTGFPVNTLLPSGLGTLITLITNRLPSYLGSSFAFIPLTATQV
      HGI:GVOIGILVAGLVAGFVKAAGKRVDAVMPAVTGAIVALLIGLNLPTAAG
      NFSSQLVATATLPAIIATVAGSGMIABLGILIGVIGWFAAITGNLSEGAADTIR
      EAAWGLGPFHKEPQLSAILLVLPVITVILIAENVGHVXAVSEMTEGDDDLGADALI
      ADGGFTLAGAGGSGTTTAAENIGVNAATRVSTAAVYVAACTAIALAFIFKFGALI
      FTTPAGVLGACILVLYGLIMLGIRIQDNKVNFPNLTMAAVALVAGIGNLTIV
      FGVTLLEGIAWAL"
ORIGIN
  Query Match 66.7%; Score 18; DB 6; Length 1335;
  Best Local Similarity 80.8%; Pred. No. 8.4e+02;
  Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 190 CAGCGCTCTTTAGGAGCGACACTG 165

RESULT 18
AX067037/c
LOCUS AX067037 1410 bp. DNA linear PAT 24-JAN-2001
DEFINITION Sequence 619 from Patent WO0100805.
ACCESSION AX067037
VERSION AX067037.1 GI:12544745
KEYWORDS
SOURCE
  Corynebacterium glutamicum
  Corynebacterium glutamicum
  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
  Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in
  membrane synthesis and membrane transport
JOURNAL Patent: WO 0100805-A 619 04-JAN-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
  source
    1. .1410
      /location/Qualifiers
      /organism="Corynebacterium glutamicum"
      /mol_type="unassigned DNA"
      /db_xref="taxon:1718"
      101. .1390
      /notes="unnamed protein product; RXN02233"
      /codon_start=1
      101. .1390
      /notes="unnamed protein product; RXN02233"
      /codon_start=1
      101. .1390
      /notes="unnamed protein product; RXN02233"
      /codon_start=1

```


located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA42170.1"
/db_xref="GI:17739559"
/translators="MLKSTQVLFASAAALLTSLAPAFALDGGDVLKKNIAAAYAQQG
EIAAGSVANGDVTGNGVTFNFAADPAKKI PVGNITLEGVEENNGVYKKNARFDNI
DYQKEVIEIKASDINYSGLDPAIDATYVMSGLFVDEAHSGPI SVSDVQKAFLEE
TSATMSVDDKNSIGPGLDASGPKADLSVTDPAIKLQALSGDMTMSGSL
VASTGVDEYAIIDDDVPLNMSFGSGVTLAFIKSMOETVKAQAANPEKADQAA
SLANGLMGQULSFVNAIEISFEDASITKRAIDYAAASQGMGTGDLAQCTLKGMAGJMAAS
LNVPELQNMVSTAINAYIDNPKNFSITAPKEKPVFFPMINGAAMGAPNTLPKMLGVTV
TANE"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDLVPPGGDDNIHPVDLKAALSERIYALALSTIMHRALPDV
RDGLKPVHRIIHAMSENGIRPNSAFKKCARIVGDVIGKPHGDCDSVDALVRVQD
FSQYYPVDDGGNGTNDGCAAYRTAEMTDVAALLGELGEDAVDFRATYNEED
EPPVLPKAPNLLANGASGTAVGMAITIPHNABELCDAAHLIKHPDATVEKLFV
IPGDLPTGGVYIESRNLIDAYKTGRGGRVRAKWTEDLGRGGYQVITIPFQVQ
KSRLIEATLLARKPLLEDVDESAEDRVILVPKRNTVDATLLMESLFLSLSLE
SRIPNNVSLGKVPKVMALNEVLSMLAHRKDVLAHRSRHLAALDRLELTGGLL
IAYLNDEIVIRIIEEEDPKQVMARSLTDNCAEALNMRNLNPKLESEPEIRKEHD
DLSEKSEIEGLASDDKQWOTVAWEIGEVKKYAKATEIGRRRTGFADAPADIAEAL
QQAMIEKPEYTVTSQKGMIRALKGHMSDTSALTFFKEGGPFLKAFPAQTITDKLLLT
GKAYATLGAIDLKPGCRGHGPIRLMVMENDODITAFVHPDPSKILLTSGGNGFV
AESVIANTRKGKIMKVSMPDEAKLAPVTDGHVAVGEMRKLALPFLSQVPEMSRG
KGVRLQYKDGSDVDKCFALADGLSWSDTAGRAFTKVGELREWLADRAATVGRTPVK
GPKRSKPGG"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDLVPPGGDDNIHPVDLKAALSERIYALALSTIMHRALPDV
RDGLKPVHRIIHAMSENGIRPNSAFKKCARIVGDVIGKPHGDCDSVDALVRVQD
FSQYYPVDDGGNGTNDGCAAYRTAEMTDVAALLGELGEDAVDFRATYNEED
EPPVLPKAPNLLANGASGTAVGMAITIPHNABELCDAAHLIKHPDATVEKLFV
IPGDLPTGGVYIESRNLIDAYKTGRGGRVRAKWTEDLGRGGYQVITIPFQVQ
KSRLIEATLLARKPLLEDVDESAEDRVILVPKRNTVDATLLMESLFLSLSLE
SRIPNNVSLGKVPKVMALNEVLSMLAHRKDVLAHRSRHLAALDRLELTGGLL
IAYLNDEIVIRIIEEEDPKQVMARSLTDNCAEALNMRNLNPKLESEPEIRKEHD
DLSEKSEIEGLASDDKQWOTVAWEIGEVKKYAKATEIGRRRTGFADAPADIAEAL
QQAMIEKPEYTVTSQKGMIRALKGHMSDTSALTFFKEGGPFLKAFPAQTITDKLLLT
GKAYATLGAIDLKPGCRGHGPIRLMVMENDODITAFVHPDPSKILLTSGGNGFV
AESVIANTRKGKIMKVSMPDEAKLAPVTDGHVAVGEMRKLALPFLSQVPEMSRG
KGVRLQYKDGSDVDKCFALADGLSWSDTAGRAFTKVGELREWLADRAATVGRTPVK
GPKRSKPGG"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDLVPPGGDDNIHPVDLKAALSERIYALALSTIMHRALPDV
RDGLKPVHRIIHAMSENGIRPNSAFKKCARIVGDVIGKPHGDCDSVDALVRVQD
FSQYYPVDDGGNGTNDGCAAYRTAEMTDVAALLGELGEDAVDFRATYNEED
EPPVLPKAPNLLANGASGTAVGMAITIPHNABELCDAAHLIKHPDATVEKLFV
IPGDLPTGGVYIESRNLIDAYKTGRGGRVRAKWTEDLGRGGYQVITIPFQVQ
KSRLIEATLLARKPLLEDVDESAEDRVILVPKRNTVDATLLMESLFLSLSLE
SRIPNNVSLGKVPKVMALNEVLSMLAHRKDVLAHRSRHLAALDRLELTGGLL
IAYLNDEIVIRIIEEEDPKQVMARSLTDNCAEALNMRNLNPKLESEPEIRKEHD
DLSEKSEIEGLASDDKQWOTVAWEIGEVKKYAKATEIGRRRTGFADAPADIAEAL
QQAMIEKPEYTVTSQKGMIRALKGHMSDTSALTFFKEGGPFLKAFPAQTITDKLLLT
GKAYATLGAIDLKPGCRGHGPIRLMVMENDODITAFVHPDPSKILLTSGGNGFV
AESVIANTRKGKIMKVSMPDEAKLAPVTDGHVAVGEMRKLALPFLSQVPEMSRG
KGVRLQYKDGSDVDKCFALADGLSWSDTAGRAFTKVGELREWLADRAATVGRTPVK
GPKRSKPGG"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDLVPPGGDDNIHPVDLKAALSERIYALALSTIMHRALPDV
RDGLKPVHRIIHAMSENGIRPNSAFKKCARIVGDVIGKPHGDCDSVDALVRVQD
FSQYYPVDDGGNGTNDGCAAYRTAEMTDVAALLGELGEDAVDFRATYNEED
EPPVLPKAPNLLANGASGTAVGMAITIPHNABELCDAAHLIKHPDATVEKLFV
IPGDLPTGGVYIESRNLIDAYKTGRGGRVRAKWTEDLGRGGYQVITIPFQVQ
KSRLIEATLLARKPLLEDVDESAEDRVILVPKRNTVDATLLMESLFLSLSLE
SRIPNNVSLGKVPKVMALNEVLSMLAHRKDVLAHRSRHLAALDRLELTGGLL
IAYLNDEIVIRIIEEEDPKQVMARSLTDNCAEALNMRNLNPKLESEPEIRKEHD
DLSEKSEIEGLASDDKQWOTVAWEIGEVKKYAKATEIGRRRTGFADAPADIAEAL
QQAMIEKPEYTVTSQKGMIRALKGHMSDTSALTFFKEGGPFLKAFPAQTITDKLLLT
GKAYATLGAIDLKPGCRGHGPIRLMVMENDODITAFVHPDPSKILLTSGGNGFV
AESVIANTRKGKIMKVSMPDEAKLAPVTDGHVAVGEMRKLALPFLSQVPEMSRG
KGVRLQYKDGSDVDKCFALADGLSWSDTAGRAFTKVGELREWLADRAATVGRTPVK
GPKRSKPGG"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDLVPPGGDDNIHPVDLKAALSERIYALALSTIMHRALPDV
RDGLKPVHRIIHAMSENGIRPNSAFKKCARIVGDVIGKPHGDCDSVDALVRVQD
FSQYYPVDDGGNGTNDGCAAYRTAEMTDVAALLGELGEDAVDFRATYNEED
EPPVLPKAPNLLANGASGTAVGMAITIPHNABELCDAAHLIKHPDATVEKLFV
IPGDLPTGGVYIESRNLIDAYKTGRGGRVRAKWTEDLGRGGYQVITIPFQVQ
KSRLIEATLLARKPLLEDVDESAEDRVILVPKRNTVDATLLMESLFLSLSLE
SRIPNNVSLGKVPKVMALNEVLSMLAHRKDVLAHRSRHLAALDRLELTGGLL
IAYLNDEIVIRIIEEEDPKQVMARSLTDNCAEALNMRNLNPKLESEPEIRKEHD
DLSEKSEIEGLASDDKQWOTVAWEIGEVKKYAKATEIGRRRTGFADAPADIAEAL
QQAMIEKPEYTVTSQKGMIRALKGHMSDTSALTFFKEGGPFLKAFPAQTITDKLLLT
GKAYATLGAIDLKPGCRGHGPIRLMVMENDODITAFVHPDPSKILLTSGGNGFV
AESVIANTRKGKIMKVSMPDEAKLAPVTDGHVAVGEMRKLALPFLSQVPEMSRG
KGVRLQYKDGSDVDKCFALADGLSWSDTAGRAFTKVGELREWLADRAATVGRTPVK
GPKRSKPGG"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDLVPPGGDDNIHPVDLKAALSERIYALALSTIMHRALPDV
RDGLKPVHRIIHAMSENGIRPNSAFKKCARIVGDVIGKPHGDCDSVDALVRVQD
FSQYYPVDDGGNGTNDGCAAYRTAEMTDVAALLGELGEDAVDFRATYNEED
EPPVLPKAPNLLANGASGTAVGMAITIPHNABELCDAAHLIKHPDATVEKLFV
IPGDLPTGGVYIESRNLIDAYKTGRGGRVRAKWTEDLGRGGYQVITIPFQVQ
KSRLIEATLLARKPLLEDVDESAEDRVILVPKRNTVDATLLMESLFLSLSLE
SRIPNNVSLGKVPKVMALNEVLSMLAHRKDVLAHRSRHLAALDRLELTGGLL
IAYLNDEIVIRIIEEEDPKQVMARSLTDNCAEALNMRNLNPKLESEPEIRKEHD
DLSEKSEIEGLASDDKQWOTVAWEIGEVKKYAKATEIGRRRTGFADAPADIAEAL
QQAMIEKPEYTVTSQKGMIRALKGHMSDTSALTFFKEGGPFLKAFPAQTITDKLLLT
GKAYATLGAIDLKPGCRGHGPIRLMVMENDODITAFVHPDPSKILLTSGGNGFV
AESVIANTRKGKIMKVSMPDEAKLAPVTDGHVAVGEMRKLALPFLSQVPEMSRG
KGVRLQYKDGSDVDKCFALADGLSWSDTAGRAFTKVGELREWLADRAATVGRTPVK
GPKRSKPGG"

gene 3077. 5329
/gene="parC"
/notes="synonym: Atu158"
3077. 5329
/gene="parC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit A"
/protein_id="AA42171.1"
/db_xref="GI:17739560"
/translators="MGKDL

```

complement (5918. 7931)
/gene="hmb"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl table=11
/product="delta-aminolevulinic acid denhydratase"
/protein_id="AAL42174.1"
/db_xref="GI:17739563"
/translation="MQDKTHLVDRITGHRBRNRKADTRRLVQENRLTVDDLTPV
FVPSGNIVEPIAMPGVNPNMSORLVEAREADGIGPAIATFPIEMELDEQGSN
ALEANNINQATIVAKVAPNGIMTVALDPTTSHGDILRGDSIVNDEIVEQVVT
AAILQAEAGSDIISPEMGRICAGIRGADLAQNGVMAYATKFSGGYIGVPYREA
ISTAGLLGDKNSYVISPANGMEAIRDAALDVEEGADMVKPGLPLDICWRLKREPI
GLPTFAYQVSGEYTIQIAAAMNGWIDGERVWMETLLCFKRAQGDGILTYFAIEAARK
AKH"
8057. 8542
/gene="Atu1162"
8057. 8542
/notes="Atu1162"
/gene="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl table=11
/product="conserved hypothetical protein"
/protein_id="AAL42175.1"
/db_xref="GI:17739564"
/translation="MQYGLVARPGRPYLPAVEPDSQPLPKRPTEILFLVFMRLVAVS
CFWFGYQWMLTGTSLGCHGRFDMNLPMRWAGSALAVFPVAALGLMUGASWGMV
WVWGAGTQIYAMTKVWPHFIFFGANTLVPVMHGLVATIIYILFLVAFWLDHRQNEERARD
p"
8774. 9286
/gene="mucs"
/notes="synonym: Atu1163"
8774. 9286
/gene="mucs"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl table=11
/product="exopolysaccharide II synthesis transcriptional
activator ExpG"
/protein_id="AAL42176.1"
/db_xref="GI:17739565"
/translation="MINSKVKPQAVVAADHTIRSLYMESLHLVERHRLLDVLIK
EFORQDQDNVAQVALLFNIGNSLITAGELSRGYLGNVSNVKKGLVDGLGLNHQ
RSVDKRSVRISLTBEGQAIETVARLYERHVHSGIEKVGIGTGFSEMNKLLQRLDR
PWDSIAYRL"
9660. 10991
/gene="Atu1164"
9660. 10991
/notes="Atu1164"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

Query Match 66.7% Score 18; DB 1; Length 11198;
Best Local Similarity 80.8%; Pred. No. 8,1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
Dh 278 CAGCCCACTTCGATCGACACCGG 253

```

| | |
|----------------|--|
| RESULT | 21 |
| AE0808044/c | |
| LOCUS | Agracacterium tumefaciens str. |
| DEFINITION | C58 circular chromosome, section 102 of 254 of the complete sequence. |
| ACCESSION | AE080804 AE007869 |
| VERSION | AE080804.1 GI:15156183 |
| SOURCE | . |
| KEYWORDS | Agracacterium tumefaciens str. C58 (Cereon) |
| DNA linear BCT | 14354 bp |
| BCT | 18-DEC-2001 |

ORGANISM Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium;
REFERENCE 1 (bases 1 to 14354)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14354)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA

FEATURES
source
1..14354
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/mol_type="genomic DNA"
/strain="C58"
/db_xref="taxon:181661"
complement(106..669)
/gene="AGR_C_2122"
complement(106..669)
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2122p"
/protein_id="AAK86950.1"
/db_xref="GI:15156184"
/translation="MPIAPASARWGTAERFAAOCPCAFVRFPQWRSPHPHP
RFPFRKATVNAIIVGADPTGETLMSECPVLPFAERQIDILEREAAWMTI
HAALERASEEVTDAFVAQSDMQPPPHDYFAAQAQQLFLLCGADPKTEGGDPEIA
GHIRNAQNIIEHYWKNDAAADIADK"
complement(623..1171)
/gene="AGR_C_2123"
complement(623..1171)
/note="hypothetical protein yung - Bacillus subtilis"
/codon_start=1
/transl_table=11
/product="AGR_C_2123p"
/protein_id="AAK86951.1"
/db_xref="GI:15156185"
/translation="MPIPERPADQVHGAPFRFRARAIPVNTSTPLSLFVLGVSG
HKTESFAGNALTETCLKRFHQDAERLYALKKSWSGETSNLRNPSKGGQS
VTALVVFQIGELAKTAVSGAEHFYNIVEGTRWDFTPAQDIPVGQDRLASRTAM
ATITPLOYAIVLSACIRKALGHS"
complement(1208..2686)
/gene="AGR_C_2125"
complement(1208..2686)
/note="probable periplasmic protein Cj0728 (imported) -
Campylobacter jejuni (strain NCTC 11168)"
/codon_start=1
/transl_table=11
/product="AGR_C_2125p"
/protein_id="AAK86952.1"
/db_xref="GI:15156186"
/translation="MPATATLKFQSFKRSQACHAATLSRQSCLOQLVRGVSTMKH
LFASCAILAFASIAHAQEPFAKLVSHAILPANTIIIPAPDAPEHLKSAKTTADR
KRAELGVTPKQDQVRLGLMPFGNQMGQSGIKAMPDGSFWSLSDNGFGSLNS
DAMLMLHLKDWADQKNALETVFLSDPMKAPPIVLESGSKYLTGADPDVESIQ
PVADGFWGEFEGPYLKFTRDGLTDVISTKAGDIEVKSPDHTLALPGNPTKQMA
FNLKSGGYEGMALSGKSLYGLLEGLYVADQVETKEDGATALRIIELDTAKKEW
TCRWLYPLAQGGGAIGDFNMDLSTALVIERDNGAGTADKACADPKKPPANCALPA
VKRYRIKLEFDKVAARAKIGYITDLLKISDPDNKRQGGNGYDMPFVTIENVR
VDATHIIVGNNDNLPPFSAGRALDKADONEFVLLEVKLLDAK"
complement(2711..3739)
/gene="AGR_C_2127"
complement(2711..3739)
/note="ADENYLATE CYCLASE 2 (ATP PYROPHOSPHATE-LYASE 2)

(ADENYLATE CYCLASE 2)"
/codon_start=1
/transl_table=11
/product="AGR_C_2127p"
/protein_id="AAK86953.1"
/db_xref="GI:15156187"
/translation="MRKLLPILDYVVLVFAVAGAVAYAFLEVGCGALICATYALPAC
APILAPERGVILPCLSRQVSALEPTPAYIAVGLIYALVFCGPGGLGTLMSLGPFG
TWKRAVHAEVQTLIFLIVCGLIIVFMRVRELLGRDIFDIDLICRVRVSEDRVIF
IDLUGSTFAETHGDKAQEFUGELIFASVAAAPVRHGGVDDIYIGRAAIIITPYHNAI
RRAACVRCVDFDIQATWEERDMWLSRFGVPRIFAIGHGPVITAEIGVDHKKITVFG
DTVNTTSRLESLSKMLGHPVLISADLAHQVLVLPKGVSEYLGHEHAVKRGQTILGVCTL
SOYQAPAA"
complement(3994..4524)
/gene="AGR_C_2127A"
complement(3994..4524)
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2127Ap"
/protein_id="AAK86954.1"
/db_xref="GI:15156188"
/translation="MKTARTLLAVLLAAAGILPSLPARAEDKVDLOQEWTTDCUKI
GKNDNGIITRITRDSRIHAYSQIYASNGCORPTQVRYEGTLDGGLDKDQNSLLFAH
TVTSITMTDNGDVVEHYNRDPKMGCGGLNDWKEILPSTMAEPAPPSPLPKPKPCS
TAPGSMTAGTFLPPFR"
complement(4606..5895)
/gene="AGR_C_2134"
complement(4606..5895)
/note="ribonuclease D PA1294 (imported) - Pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_C_2134p"
/protein_id="AAK86955.1"
/db_xref="GI:15156189"
/translation="MNRSPTRKRTILEAGKAGFVLTCAQRESKTDNFHWPFAAMIE
TTAALAACTELAKSEFIDTETFEFTFWPELCLVQMASPTLEVLPFLAKGIDLIT
PKFELMANPVVFAHAAQDTEIIVHGLGLIPHPIDFQVAAVCGFGDSISYDOLV
OKIKVQIDKSSRFTDMSRPLETEKOLDYALADVTHLRDVLVLSKLAQLEREGSLMT
EKNILLESRTYDMHPDDANLRLKSRLEKTEHLAIKLVAAAREAREARNVPSRVL
KDGIFEIACQCPDAELSRUKTIIPGWSASGTATITETVNAALALPKEENPKAPR
HSHAPESGAVALLEKVLKLTADKHGAAKVIANSDLLDKTASEGENATVAALTQWR
RELFGVALKLINGEVALRFAGKKVEAVEVEAAASES"
complement(5855..7144)
/gene="AGR_C_2135"
complement(5855..7144)
/note="hypothetical protein PA3355 (imported) -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_C_2135p"
/protein_id="AAK86956.1"
/db_xref="GI:15156190"
/translation="MSQIRPLLVLTAGILIGNGLOQTYIALRGLSGSGFASITGL
ISAYSIGALGICISVTQMRRIIGHRTFATMAVAASAASIAMPLYLHFLFWMLMRFV
IGIAVACLPAAESWINAQVNTSNRARTLSIVRFVLDGSLVTLAQYIIPAGIDGPVVF
SLTAMALTLSLVPSIADKSNTPPKQIPFNLFAVWRISPLAVGCVAVGLSMAAFRN
MGPIYAFQIGMSVTATFVSGAGILGVLGVSDRYDRRIIILATTAGSVTVG
LELFAQGTDESNIIIGVFGNAPALPLSLASGNDHAREGHEHAMAGLLPFKSV
GATVGPLLASLLIDQFGPKALSYTHAIOIYIATYLYRLQVRGVFVEARNRFRSL
LRTSAVFGKLAAPTPPKADHPKKDSEFEPPEKNDP"
7325..9112
/gene="AGR_C_2136"
7325..9112
/note="AGR_C_2136"
/note="(AL392189) aspartyl-tRNA synthetase"
/codon_start=1
/transl_table=11
/product="AGR_C_2136p"

/note="Sequence from overlapping clone RP11-311B18
(AL139818). Assembly confirmed by restriction digest."

misc_feature 56166..56252

/note="match: STS: Em:A0048351"

misc_feature 57887..58177

/note="match: STS: Em:G50546"

ORIGIN

Query Match 66.7%; Score 18; DB 9; Length 84615;
Best Local Similarity 80.8%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26

DB 29570 CACCTGTGCTTCAGGATCCTCACCTG 29545

RESULT 23
AL954650 86257 bp DNA linear PRI 26-FEB-2003
LOCUS Human DNA sequence from clone RP11-523M19 on chromosome 1, complete
DEFINITION sequence.

ACCESSION AL954650

VERSION AL954650.8 GI:28564367

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 86257)

Direct Submission

Submitted (26-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 25, 2003 this sequence version replaced gi:28208053.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

RP11-523M19 is from the library RP11-11.2 constructed by the group

of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

Location/Qualifiers

1..86257

/organism="Homo sapiens"

/mol_type="genomic DNA"

FEATURES

Source

1..109889

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="11"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-523M19"

/clone_lib="RPCI-11.2"

ORIGIN

Query Match 66.7%; Score 18; DB 9; Length 86257;
Best Local Similarity 80.8%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26

DB 5930 CACCGCGCTTCAGGATCCACACCTG 5955

RESULT 24

AL954830

LOCUS

DEFINITION

Mouse DNA sequence from clone RP23-12A10 on chromosome 11, complete

sequence.

AL954830 109889 bp DNA linear ROD 13-DEC-2002

VERSION AL954830.7 GI:26985489

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 109889)

Direct Submission

Submitted (10-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Dec 16, 2002 this sequence version replaced gi:26788337.

Sequence from the Mouse Genome Sequencing Consortium whole genome

shotgun may have been used to confirm this sequence. Sequence data

from the whole genome shotgun alone has only been used where it has

a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-12A10 is

from the RP11-23 Mouse PAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

Location/Qualifiers

1..109889

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="11"

```
AC122465.4 GI:26024143
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 142340)
The sequence of Mus musculus BAC clone RP24-304I22
Unpublished (2001)
2 (bases 1 to 142340)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 142340)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 142340)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (13-NOV-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 142340)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (04-DEC-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 142340)
Wilson, R.
Direct Submission
Submitted (13-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 4, 2002 this sequence version replaced gi:24943033.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
-----
Center project name: M_BB0304122
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. 142340

FEATURES
source
```

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/map="13"
/clone="RP24-304122"
/clone_lib="RPC1-24"
350..419
/rpt_family="ID"
429..535
/rpt_family="Alu"
941..975
/rpt_family="rRNA-Phe-TTY"
947..1128
/rpt_family="B4"
2257..2449
/rpt_family="B2"
2468..2578
/rpt_family="L1"
2648..2788
/rpt_family="Alu"
2889..3019
/rpt_family="Alu"
4711..4793
/rpt_family="ID"
4794..4831
/rpt_family="B4"
4869..4984
/rpt_family="Alu"
5046..5346
/rpt_family="L1"
5746..5836
/rpt_family="Alu"
6301..6337
/rpt_family="CR1"
6651..6782
/rpt_family="Alu"
6874..7003
/rpt_family="MaLR"
7472..7664
/rpt_family="B2"
7713..7826
/rpt_family="MIR"
7903..8227
/rpt_family="MaLR"
8404..8648
/rpt_family="B4"
9311..9449
/rpt_family="Alu"
9573..9673
/note="Sequence derived from one plasmid subclone."
10586..10755
/rpt_family="L1"
10843..10937
/rpt_family="ID"
11030..11146
/rpt_family="L1"
11147..11508
/rpt_family="MaLR"
11509..11602
/rpt_family="L1"
12000..12030
/rpt_family="B2"
12031..12189
/rpt_family="B4"
12186..12396
/rpt_family="MaLR"
12714..12878
/rpt_family="ERV1"
12879..13001
/rpt_family="Alu"
13139..13390
/rpt_family="ERV1"

```

```

repeat_region 13391..13574
/rpt_family="B2"
repeat_region 13575..13689
/rpt_family="ERV1"
repeat_region 15299..15355
/rpt_family="ID"
repeat_region 15356..15459
/rpt_family="B4"
repeat_region 16335..16412
/rpt_family="B4"
repeat_region 16413..16556
/rpt_family="Alu"
repeat_region 16560..16705
/rpt_family="Alu"
repeat_region 16706..16732
/rpt_family="BC1_XM"
repeat_region 17753..17785
/rpt_family="B2"
repeat_region 17892..17939
/rpt_family="ERV1"
repeat_region 18684..18772
/rpt_family="Alu"
repeat_region 18826..19103
/rpt_family="B4"
repeat_region 18894..19950
/rpt_family="L1"
repeat_region 19964..20066
/rpt_family="Alu"
repeat_region 20631..20971
/rpt_family="MaLR"
repeat_region 20976..21471
/rpt_family="L1"
repeat_region 23096..23124
/rpt_family="B4"
repeat_region 23125..23250
/rpt_family="Alu"
repeat_region 23262..23724
/rpt_family="L1"
repeat_region 23732..24165
/rpt_family="RMR10A"
repeat_region 24219..24298
/rpt_family="ID"
repeat_region 24316..24385
/rpt_family="B4"
repeat_region 24658..24768
/rpt_family="Alu"
repeat_region 24895..24947
/rpt_family="ERV1"
repeat_region 25277..25504
/rpt_family="B4"

```

Query Match 66.7%; Score 18; DB 10; Length 142340;

Best Local Similarity 80.8%; Pred.No. 7.7e+02;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTG 26

Db 63795 CACCCGCGCTTAGGCTCCACATG 63820

RESULT 28

AC106319/c

LOCUS

AC106319 Rattus norvegicus clone CH230-239A7, WORKING DRAFT SEQUENCE.

DEFINITION

AC106319

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 184424)

AC106319 184424 bp DNA linear HTG 13-MAY-2003

Rattus norvegicus clone CH230-239A7, WORKING DRAFT SEQUENCE.

AC106319

AC106319.4 GI:30581450

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

| | |
|--|--|
| AUTHORS | Muzny, D. Marie., Metzker, M. Lee., Abranzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, C., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemeleh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, R., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A. |
| | Direct Submission |
| | Unpublished |
| | 2 (bases 1 to 184424) |
| | Worley, K. C. |
| | Direct Submission |
| | Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| | 3 (bases 1 to 184424) |
| | Rat Genome Sequencing Consortium. |
| | Direct Submission |
| Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA | |
| On May 13, 2003 this sequence version replaced gi:23608511. | |
| The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole | |
| genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table. | |
| ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GKPP Center clone name: CH230-239A7 ----- Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 178868 bases at least Q40 Consensus quality: 180315 bases at least Q30 Consensus quality: 181270 bases at least Q20 Estimated insert size: 189201; sum-of-contigs estimation Quality coverage: 8x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank Draft data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 1 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * provided by the submittor. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved. * 1 184424: contig of 184424 bp in length. Location/Qualifiers 1. 184424 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-239A7" /notes="clone boundary 1625. 2457 clone_end:Sp6 site:EcoRI end sequence:B2103095" complement(179035..179868) /notes="clone boundary clone_end:T7 site:EcoRI end sequence:B2103094" | |
| FEATURES | |
| source | |
| misc_feature | |
| misc_feature | |
| ORIGIN | |
| Query Match 66.7%; Score 18; DB 2; Length 184424; Best Local Similarity 80.8%; Pred. No. 7.7e+02; Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0; | |
| Oy 2 ACCGCGCTTAGGATCGACACCTGA 27 | |
| Db 28044 AGCGAGCTTAGGTTTGAAACCTGA 28019 | |
| RESULT 29 | |
| AC084412/c | |
| LOCUS | |
| DEFINITION Mus musculus clone RP23-174G20, WORKING DRAFT SEQUENCE, 32 | |
| AC084412 206112 bp DNA linear HTG 01-NOV-2000 | |
| unordered pieces. | |
| AC084412 | |
| AC084412.1 GI:11067261 | |
| HTG; HTGS PHASE1; HTGS DRAFT. | |
| KEYWORDS | |
| SOURCE | |
| ORGANISM | |
| Mus musculus | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | |
| 1 (bases 1 to 206112) | |
| REFERENCE | |
| AUTHORS | |
| McCombie, W. R., Baker, J. P., Bahret, A., Bal, H., Balija, V., Dedhia, N. N., de la Bastide, M., Huang, E. N., King, L., Kirchoff, K. A., | |

Miller, B., Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Rodriguez, S., Santos, L., Shah, R.S., Spiegel, L.A., Toch, K., Vil, M.D. and Zucaverni, T.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Mouse Genomic Sequence
Unpublished
2 (bases 1 to 206112)
McCombie, W.R.

Direct Submission

Submitted (01-NOV-2000) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

COMMENT

----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory

Center code: CSHL

Web site: <http://www.cshl.org/genseq>

Contact: mccombie@cshl.org

----- Project Information

Center project name: RP23-174G20

Center clone name: RP23-174G20

*** NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 17043: contig of 17043 bp in length
1 17214: gap of unknown length
17215 32036: contig of 14822 bp in length
32037 32207: gap of unknown length
32208 46766: contig of 14559 bp in length
46767 46937: gap of unknown length
46938 57574: contig of 10637 bp in length
57575 57745: gap of unknown length
57746 68166: contig of 10421 bp in length
68167 68338: gap of unknown length
68339 77389: contig of 9052 bp in length
77390 77560: gap of unknown length
77561 86460: contig of 8900 bp in length
86461 86630: gap of unknown length
86631 94462: contig of 7832 bp in length
94463 102252: gap of unknown length
102253 102422: gap of unknown length
102423 109611: contig of 7189 bp in length
109612 109781: gap of unknown length
109782 116602: contig of 6821 bp in length
116603 116773: gap of unknown length
116773 123357: contig of 6585 bp in length
123358 123527: gap of unknown length
123528 123808: contig of 6281 bp in length
123809 129978: gap of unknown length
129979 135993: contig of 6015 bp in length
135994 136164: gap of unknown length
136164 141952: contig of 5788 bp in length
141952 142122: gap of unknown length
142122 147858: contig of 5737 bp in length
147859 148028: gap of unknown length
148029 153410: contig of 5382 bp in length
153411 153580: gap of unknown length
153581 153783: contig of 5203 bp in length
153784 159953: gap of unknown length
159954 163793: contig of 4840 bp in length
163794 163794: gap of unknown length
163794 169797: contig of 4834 bp in length
169798 169798: gap of unknown length
169798 173706: contig of 4739 bp in length
173707 173876: gap of unknown length
173877 177928: contig of 4052 bp in length
177929 178098: gap of unknown length

* 178099 181857: contig of 3759 bp in length
* 181858 182027: gap of unknown length
* 182028 184932: contig of 2905 bp in length
* 184933 185102: gap of unknown length
* 185103 187993: contig of 2891 bp in length
* 187994 188163: gap of unknown length
* 188164 191028: contig of 2865 bp in length
* 191029 191199: gap of unknown length
* 191199 193997: contig of 2798 bp in length
* 193997 194167: gap of unknown length
* 194167 196954: contig of 2787 bp in length
* 196954 197124: gap of unknown length
* 197124 199647: contig of 2524 bp in length
* 199648 199817: gap of unknown length
* 199818 202136: contig of 2318 bp in length
* 202136 202306: gap of unknown length
* 202306 204590: contig of 2284 bp in length
* 204590 204759: gap of unknown length
* 204759 206112: contig of 1353 bp in length.

FEATURES

Location/Qualifiers

1..206112
source
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-174G20"

ORIGIN

Query Match 66.7%; Score 18; DB 2; Length 206112;
Best Local Similarity 80.8%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27

Db 74245 ACCCGCGCTTTCTGATCTACACGAGA 74220

RESULT 30

AC099282/c

LOCUS Rattus norvegicus clone CH230-130019, *** SEQUENCING IN PROGRESS
DEFINITION Rattus norvegicus 213462 bp DNA linear HTG 21-SEP-2002
*** 4 unordered pieces.

AC099282

AC099282.10 GI:23265682

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 213462)

REFERENCE

AUTHORS

Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draber, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

```

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasceriak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, Z., Song, X.-Z., Sorrelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R.A., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 213462)
Worley, K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 213462)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21795359.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKCN
Center clone name: CH230-130019
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208516 bases at least Q40
Consensus quality: 209366 bases at least Q30
Consensus quality: 210059 bases at least Q20
Estimated insert size: 227892; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/dccs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 166582: contig of 166582 bp in length
* 166583 166682: gap of unknown length
* 166683 168153: contig of 1471 bp in length
* 168154 168253: gap of unknown length
* 168254 183417: contig of 15164 bp in length
* 183418 183517: gap of unknown length
* 183518 213462: contig of 29945 bp in length.
* Location/Qualifiers
* 1..213462
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-130019"
* complement(163258..164086)
* /note="clone boundary"
* clone_end:T7
* site:EcoRI
* end_sequence:BH342154"

ORIGIN
Query Match 66.7%; Score 18; DB 2; Length 213462;
Best Local Similarity 80.8%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
Db 56485 CACCCGCGCTTTAGGATCGACACCTG 56460

Search completed: June 20, 2004, 11:42:32
Job time : 760.143 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 155.143 seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-12
Perfect score: 27
Sequence: 1 caccgcgcgttaggcacacctga 27

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : N Geneseq 29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------|
| C 1 | 27 | 100.0 | 110000 | 4 | AAI199682_25 |
| C 2 | 27 | 100.0 | 110000 | 4 | AAI199682_26 |
| C 3 | 27 | 100.0 | 110000 | 4 | AAI199683_25 |
| C 4 | 27 | 100.0 | 110000 | 4 | AAI199683_26 |
| C 5 | 18.6 | 68.9 | 2000 | 6 | ABZ16792 |
| C 6 | 18 | 66.7 | 1287 | 5 | AAH67093 |
| C 7 | 18 | 66.7 | 1335 | 4 | AAH68053 |
| C 8 | 18 | 66.7 | 1410 | 4 | AAH68052 |
| C 9 | 18 | 66.7 | 34980 | 5 | AAH68030 |
| C 10 | 17.6 | 65.2 | 766 | 4 | AAH68030 |
| C 11 | 17.6 | 65.2 | 771 | 4 | AAH68030 |
| C 12 | 17.6 | 65.2 | 894 | 7 | AAH68030 |
| C 13 | 17.6 | 65.2 | 933 | 7 | AAH68030 |
| C 14 | 17.6 | 65.2 | 1195 | 5 | AAH68030 |
| C 15 | 17.6 | 65.2 | 4919 | 6 | AAH68030 |
| C 16 | 17.4 | 64.4 | 1789 | 7 | AAH68030 |
| C 17 | 17.4 | 64.4 | 2283 | 7 | AAH68030 |
| C 18 | 17.4 | 64.4 | 2283 | 7 | AAH68030 |
| C 19 | 17.4 | 64.4 | 2937 | 4 | AAH68030 |
| C 20 | 17.4 | 64.4 | 5484 | 4 | AAH68030 |
| C 21 | 17.4 | 64.4 | 110000 | 4 | AAI199682_40 |
| C 22 | 17.4 | 64.4 | 110000 | 4 | AAI199683_40 |
| C 23 | 17.2 | 63.7 | 999 | 6 | ABK74646 |

| | | | | | |
|------|------|------|-------|---|-----------|
| 24 | 17.2 | 63.7 | 1257 | 7 | ACA36049 |
| C 25 | 17 | 63.0 | 279 | 4 | AAK44543 |
| C 26 | 17 | 63.0 | 279 | 4 | ABS44201 |
| C 27 | 17 | 63.0 | 279 | 6 | ABS18780 |
| C 28 | 17 | 63.0 | 349 | 6 | ABS69598 |
| C 29 | 17 | 63.0 | 463 | 4 | AAK31460 |
| C 30 | 17 | 63.0 | 463 | 4 | ABS31142 |
| C 31 | 17 | 63.0 | 463 | 6 | ABS06214 |
| C 32 | 17 | 63.0 | 618 | 3 | AAA44965 |
| C 33 | 17 | 63.0 | 897 | 7 | ACA28624 |
| C 34 | 17 | 63.0 | 1485 | 7 | ACA20990 |
| C 35 | 17 | 63.0 | 1674 | 7 | ACA36278 |
| C 36 | 17 | 63.0 | 1675 | 7 | ACA32299 |
| C 37 | 17 | 63.0 | 1701 | 2 | AAQ65617 |
| C 38 | 17 | 63.0 | 1701 | 2 | AAV64813 |
| C 39 | 17 | 63.0 | 1701 | 2 | AAZ22736 |
| C 40 | 17 | 63.0 | 1701 | 3 | AAZ33270 |
| C 41 | 17 | 63.0 | 1701 | 3 | AAZ95673 |
| C 42 | 17 | 63.0 | 1701 | 3 | AAZ46281 |
| C 43 | 17 | 63.0 | 1701 | 3 | AAZ37820 |
| C 44 | 17 | 63.0 | 12409 | 4 | AAZ30238 |
| C 45 | 17 | 63.0 | 12409 | 7 | ACD01493 |
| C 46 | 17 | 63.0 | 12838 | 2 | AAZ20509 |
| C 47 | 17 | 63.0 | 17736 | 4 | ABL28554 |
| C 48 | 16.8 | 62.2 | 8163 | 4 | AAK84484 |
| C 49 | 16.8 | 62.2 | 8163 | 4 | AAK05749 |
| C 50 | 16.6 | 61.5 | 467 | 6 | ABL93494 |
| C 51 | 16.6 | 61.5 | 546 | 6 | ABZ12882 |
| C 52 | 16.6 | 61.5 | 587 | 2 | AAI19928 |
| C 53 | 16.6 | 61.5 | 920 | 7 | ABZ51070 |
| C 54 | 16.6 | 61.5 | 939 | 7 | ACA40224 |
| C 55 | 16.6 | 61.5 | 1357 | 5 | AAZ85511 |
| C 56 | 16.6 | 61.5 | 1467 | 3 | AAZ32779 |
| C 57 | 16.6 | 61.5 | 1587 | 7 | ABX78350 |
| C 58 | 16.6 | 61.5 | 1675 | 7 | ACA49450 |
| C 59 | 16.6 | 61.5 | 1677 | 7 | ACA51446 |
| C 60 | 16.6 | 61.5 | 1712 | 5 | AAZ77240 |
| C 61 | 16.6 | 61.5 | 1901 | 3 | AAZ98864 |
| C 62 | 16.6 | 61.5 | 1995 | 7 | ACA23852 |
| C 63 | 16.6 | 61.5 | 2218 | 5 | AAZ74300 |
| C 64 | 16.6 | 61.5 | 2575 | 5 | AAZ71248 |
| C 65 | 16.6 | 61.5 | 3164 | 4 | ABL25304 |
| C 66 | 16.6 | 61.5 | 3211 | 5 | AAZ77492 |
| C 67 | 16.6 | 61.5 | 5927 | 4 | ABL6006 |
| C 68 | 16.6 | 61.5 | 5927 | 4 | AAZ57148 |
| C 69 | 16.6 | 61.5 | 5927 | 9 | ADC35863 |
| C 70 | 16.4 | 60.7 | 100 | 7 | ACD80473 |
| C 71 | 16.4 | 60.7 | 102 | 2 | AAQ70234 |
| C 72 | 16.4 | 60.7 | 258 | 3 | AAI15651 |
| C 73 | 16.4 | 60.7 | 318 | 2 | AAQ12499 |
| C 74 | 16.4 | 60.7 | 423 | 4 | AAI193517 |
| C 75 | 16.4 | 60.7 | 427 | 2 | AAQ70220 |
| C 76 | 16.4 | 60.7 | 438 | 3 | AAI15656 |
| C 77 | 16.4 | 60.7 | 555 | 2 | AAQ03732 |
| C 78 | 16.4 | 60.7 | 555 | 2 | AAQ02288 |
| C 79 | 16.4 | 60.7 | 595 | 4 | AAI19308 |
| C 80 | 16.4 | 60.7 | 595 | 4 | ABA31457 |
| C 81 | 16.4 | 60.7 | 595 | 4 | AAK38507 |
| C 82 | 16.4 | 60.7 | 595 | 4 | AAI12784 |
| C 83 | 16.4 | 60.7 | 633 | 1 | AAZ92729 |
| C 84 | 16.4 | 60.7 | 698 | 2 | AAQ10159 |
| C 85 | 16.4 | 60.7 | 723 | 9 | ADC92366 |
| C 86 | 16.4 | 60.7 | 735 | 6 | ABK74283 |
| C 87 | 16.4 | 60.7 | 823 | 2 | AAQ13323 |
| C 88 | 16.4 | 60.7 | 902 | 2 | AAQ81202 |
| C 89 | 16.4 | 60.7 | 951 | 2 | AAQ11967 |
| C 90 | 16.4 | 60.7 | 975 | 7 | ACA51639 |
| C 91 | 16.4 | 60.7 | 1004 | 7 | ACA49007 |
| C 92 | 16.4 | 60.7 | 1011 | 2 | ACA31854 |
| C 93 | 16.4 | 60.7 | 1153 | 2 | AAQ92410 |
| C 94 | 16.4 | 60.7 | 1188 | 7 | ACA44613 |
| C 95 | 16.4 | 60.7 | 1368 | 2 | AAV04530 |
| C 96 | 16.4 | 60.7 | 1742 | 5 | AAZ89852 |

| | |
|-----------|------------|
| ACA36049 | Prokaryot |
| AAK44543 | Human bon |
| ABS44201 | Human liv |
| ABS18780 | Human gen |
| ABS69598 | Novel mur |
| AAK31460 | Human bon |
| ABS31142 | Human liv |
| ABS06214 | Human gen |
| AAA44965 | Human sec |
| ACA28624 | Prokaryot |
| ACA20990 | Prokaryot |
| ACA36278 | Prokaryot |
| ACA32299 | Prokaryot |
| AAQ65617 | Human zon |
| AAV64813 | Human ZPB |
| AAZ22736 | Human zon |
| AAZ33270 | Human zon |
| AAZ95673 | Human zon |
| AAZ46281 | Human ooc |
| AAZ37820 | Human zon |
| AAZ30238 | DNA encod |
| ACD01493 | Human gen |
| AAZ20509 | Polynucle |
| ABL28554 | Drosophil |
| AAK84484 | Human imm |
| AAK05749 | Human rep |
| ABL93494 | Arabidops |
| ABZ12882 | Arabidops |
| AAI19928 | Lily male |
| ABZ51070 | Aspergill |
| ACA40224 | Prokaryot |
| AAZ85511 | DNA encod |
| AAZ32779 | Arabidops |
| ABX78350 | Wheat str |
| ACA49450 | Prokaryot |
| ACA51446 | Prokaryot |
| AAZ77240 | DNA encod |
| AAZ98864 | Human sec |
| ACA23852 | Prokaryot |
| AAZ74300 | DNA encod |
| AAZ71248 | DNA encod |
| ABL25304 | Drosophil |
| AAZ77492 | DNA encod |
| ABL6006 | Drosophil |
| AAZ57148 | DNA encod |
| ADC35863 | Drosophil |
| ACD80473 | E. coli K |
| AAQ70234 | HTLV-1 3A3 |
| AAI15651 | HTLV gp21 |
| AAQ12499 | Nucleotid |
| AAI193517 | Human pol |
| AAQ70220 | HTLV p21 |
| AAI15656 | HTLV anti |
| AAQ03732 | Gene enco |
| AAQ02288 | Human T-c |
| AAI19308 | Probe #92 |
| ABA31457 | Probe #99 |
| AAK38507 | Human bon |
| AAI12784 | Human bra |
| AAZ92729 | HIV-1 - h |
| AAQ10159 | Sequence |
| ADC92366 | E. faeciu |
| ABK74283 | Bacillus |
| AAQ13323 | HTLV-1 en |
| AAQ81202 | Cloned al |
| AAQ11967 | HIV-1/HTL |
| ACA51639 | Prokaryot |
| ACA49007 | Prokaryot |
| ACA31854 | Prokaryot |
| AAQ92410 | Escherich |
| ACA44613 | Prokaryot |
| AAV04530 | HTLV-I 5' |
| AAZ89852 | DNA encod |

| | | | | | | | | | | | | | | | |
|-----|------|------|--------|---|-----------|-----------|------------------------|-------|------|------|-------|---|----------|-----------|-----------------|
| 97 | 16.4 | 60.7 | 1797 | 2 | AAG11368 | Aac11968 | HIV-2/HIV | c 170 | 16 | 59.3 | 2307 | 7 | ACA57281 | Aca57281 | Human adi |
| 98 | 16.4 | 60.7 | 2095 | 4 | ABLI13403 | Abli13403 | Drosophila | c 171 | 16 | 59.3 | 2445 | 9 | ADB69902 | Adb69902 | C. neoformans |
| 99 | 16.4 | 60.7 | 2165 | 2 | AAT59702 | Aat59702 | Endotoxin | c 172 | 16 | 59.3 | 2466 | 9 | ADB69900 | Adb69900 | C. neoformans |
| 100 | 16.4 | 60.7 | 2220 | 1 | AAN91238 | Aan91238 | Bi antigen | c 173 | 16 | 59.3 | 2553 | 7 | ACA43329 | Aca43329 | Prokaryote |
| 101 | 16.4 | 60.7 | 2331 | 6 | ACA24911 | Aca24911 | Prokaryote | c 174 | 16 | 59.3 | 2871 | 7 | ACA25101 | Aca25101 | Prokaryote |
| 102 | 16.4 | 60.7 | 2565 | 6 | ABT10715 | Abt10715 | Human breast | c 175 | 16 | 59.3 | 2880 | 2 | AAT15539 | Aat15539 | C. neoformans |
| 103 | 16.4 | 60.7 | 2576 | 9 | AD224910 | Ad224910 | DNA encoding | c 176 | 16 | 59.3 | 2882 | 9 | ADB69539 | Adb69539 | C. neoformans |
| 104 | 16.4 | 60.7 | 2665 | 6 | ABN97212 | Abn97212 | Gene #371 | c 177 | 16 | 59.3 | 2931 | 4 | AAD18505 | Aad18505 | Aspergillus |
| 105 | 16.4 | 60.7 | 2711 | 9 | AD553725 | Ad553725 | Human protein | c 178 | 16 | 59.3 | 2954 | 9 | ADB69541 | Adb69541 | C. neoformans |
| 106 | 16.4 | 60.7 | 2761 | 9 | AD553725 | Ad553725 | Farnesyl | c 179 | 16 | 59.3 | 3241 | 7 | ADB69541 | Adb69541 | C. neoformans |
| 107 | 16.4 | 60.7 | 2865 | 4 | AAK52354 | Aak52354 | Human poliovirus | c 180 | 16 | 59.3 | 3891 | 7 | ACA54433 | Aca54433 | Prokaryote |
| 108 | 16.4 | 60.7 | 2919 | 4 | ABLI12405 | Abli12405 | Drosophila | c 181 | 16 | 59.3 | 4021 | 8 | ADC19184 | Adc19184 | E. coli O157 |
| 109 | 16.4 | 60.7 | 2951 | 4 | AAK53338 | Aak53338 | Human poliovirus | c 182 | 16 | 59.3 | 4021 | 8 | ADC19184 | Adc19184 | E. coli O157 |
| 110 | 16.4 | 60.7 | 3054 | 5 | AAK585672 | Aak585672 | DNA encoding | c 183 | 16 | 59.3 | 4156 | 2 | AAT19067 | Aat19067 | Human endotoxin |
| 111 | 16.4 | 60.7 | 3244 | 4 | ABLI17295 | Abli17295 | Drosophila | c 184 | 16 | 59.3 | 4156 | 2 | AAT19067 | Aat19067 | Human endotoxin |
| 112 | 16.4 | 60.7 | 4303 | 5 | AAK585519 | Aak585519 | DNA encoding | c 185 | 16 | 59.3 | 4156 | 3 | AAK5164 | Aak5164 | Human endotoxin |
| 113 | 16.4 | 60.7 | 4692 | 4 | ABLI13402 | Abli13402 | Drosophila | c 186 | 16 | 59.3 | 4156 | 3 | AAK5164 | Aak5164 | Human endotoxin |
| 114 | 16.4 | 60.7 | 5262 | 4 | ABLI12404 | Abli12404 | Drosophila | c 187 | 16 | 59.3 | 4156 | 3 | AAK5164 | Aak5164 | Human endotoxin |
| 115 | 16.4 | 60.7 | 5813 | 4 | AAH57490 | Aah57490 | Human liver | c 188 | 16 | 59.3 | 4156 | 7 | ACA62293 | Aca62293 | Human endotoxin |
| 116 | 16.4 | 60.7 | 7095 | 4 | ABLI17294 | Abli17294 | Drosophila | c 189 | 16 | 59.3 | 4156 | 7 | ACA62293 | Aca62293 | Human endotoxin |
| 117 | 16.4 | 60.7 | 7287 | 4 | AAH57428 | Aah57428 | Human intestine | c 190 | 16 | 59.3 | 4738 | 9 | ADB69178 | Adb69178 | C. neoformans |
| 118 | 16.4 | 60.7 | 9045 | 2 | AAK42902 | Aak42902 | DNA sequence | c 191 | 16 | 59.3 | 4954 | 9 | ADB69178 | Adb69178 | C. neoformans |
| 119 | 16.4 | 60.7 | 9047 | 1 | AAK40080 | Aak40080 | Sequence | c 192 | 16 | 59.3 | 4954 | 9 | ADB69178 | Adb69178 | C. neoformans |
| 120 | 16.4 | 60.7 | 13852 | 5 | ABAI17744 | Abai17744 | Human nerve | c 193 | 16 | 59.3 | 13175 | 4 | ABL02816 | Abli02816 | Drosophila |
| 121 | 16.4 | 60.7 | 13852 | 5 | ABAI17744 | Abai17744 | Human nerve | c 194 | 16 | 59.3 | 13175 | 4 | ABL02816 | Abli02816 | Drosophila |
| 122 | 16.4 | 60.7 | 13852 | 5 | ABAI17744 | Abai17744 | Human nerve | c 195 | 16 | 59.3 | 13175 | 4 | ABL02816 | Abli02816 | Drosophila |
| 123 | 16.4 | 60.7 | 13852 | 5 | ABAI17744 | Abai17744 | Human nerve | c 196 | 16 | 59.3 | 13175 | 4 | ABL02816 | Abli02816 | Drosophila |
| 124 | 16.4 | 60.7 | 18776 | 4 | AAK59508 | Aak59508 | Propionibacterium | c 197 | 16 | 59.3 | 13611 | 7 | ABZ96982 | Abz96982 | Human endotoxin |
| 125 | 16.4 | 60.7 | 23078 | 4 | AAK59508 | Aak59508 | Propionibacterium | c 198 | 16 | 59.3 | 13611 | 7 | ABZ96982 | Abz96982 | Human endotoxin |
| 126 | 16.4 | 60.7 | 23078 | 4 | AAK59508 | Aak59508 | Propionibacterium | c 199 | 16 | 59.3 | 13611 | 7 | ABZ96982 | Abz96982 | Human endotoxin |
| 127 | 16.4 | 60.7 | 65011 | 6 | AD579015 | Ad579015 | Mouse Mef | c 200 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 128 | 16.4 | 60.7 | 96599 | 9 | AD579015 | Ad579015 | Mouse Mef | c 201 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 129 | 16.4 | 60.7 | 96599 | 9 | AD579015 | Ad579015 | Mouse Mef | c 202 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 130 | 16.4 | 60.7 | 96599 | 9 | AD579015 | Ad579015 | Mouse Mef | c 203 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 131 | 16.4 | 60.7 | 110000 | 2 | AAK42063 | Aak42063 | Continuation (6 of 42) | c 204 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 132 | 16.4 | 60.7 | 110000 | 4 | AAK42063 | Aak42063 | Continuation (6 of 42) | c 205 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 133 | 16.4 | 60.7 | 110000 | 4 | AAK42063 | Aak42063 | Continuation (6 of 42) | c 206 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 134 | 16.4 | 60.7 | 319608 | 5 | AAK593301 | Aak593301 | Human chromosome | c 207 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 135 | 16.4 | 60.7 | 319608 | 5 | AAK593301 | Aak593301 | Human chromosome | c 208 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 136 | 16.2 | 60.0 | 435 | 6 | ABN23757 | Abn23757 | Human chromosome | c 209 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 137 | 16.2 | 60.0 | 1156 | 6 | ABN23757 | Abn23757 | Human chromosome | c 210 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 138 | 16.2 | 60.0 | 1554 | 6 | ABN23757 | Abn23757 | Human chromosome | c 211 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 139 | 16.2 | 60.0 | 1554 | 6 | ABN23757 | Abn23757 | Human chromosome | c 212 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 140 | 16.2 | 60.0 | 1605 | 9 | AAK585595 | Aak585595 | DNA encoding | c 213 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 141 | 16.2 | 60.0 | 2239 | 5 | AAK585595 | Aak585595 | DNA encoding | c 214 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 142 | 16.2 | 60.0 | 2784 | 6 | ABO70687 | Abg70687 | Listeria | c 215 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 143 | 16.2 | 60.0 | 3128 | 9 | AD52465 | Ad52465 | Primary rat | c 216 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 144 | 16.2 | 60.0 | 3128 | 9 | AD52465 | Ad52465 | Primary rat | c 217 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 145 | 16.2 | 60.0 | 5821 | 2 | AAK585686 | Aak585686 | DNA encoding | c 218 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 146 | 16.2 | 60.0 | 8261 | 4 | AAK64961 | Aak64961 | Human chromosome | c 219 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 147 | 16.2 | 60.0 | 12905 | 3 | AAK299101 | Aak299101 | S. fradiae | c 220 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 148 | 16.2 | 60.0 | 12905 | 3 | AAK299101 | Aak299101 | S. fradiae | c 221 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 149 | 16.2 | 60.0 | 14070 | 4 | AAK70465 | Aak70465 | Human chromosome | c 222 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 150 | 16.2 | 60.0 | 18060 | 4 | ACI24467 | Aci24467 | Human chromosome | c 223 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 151 | 16.2 | 60.0 | 25 | 8 | ACI24467 | Aci24467 | Human chromosome | c 224 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 152 | 16.2 | 60.0 | 25 | 8 | ACI24467 | Aci24467 | Human chromosome | c 225 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 153 | 16.2 | 60.0 | 516 | 5 | ABV05086 | Abv05086 | Human chromosome | c 226 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 154 | 16.2 | 60.0 | 523 | 6 | ABQ51047 | Abq51047 | Oligonucleotide | c 227 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 155 | 16.2 | 60.0 | 523 | 6 | ABQ51046 | Abq51046 | Oligonucleotide | c 228 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 156 | 16.2 | 60.0 | 576 | 6 | ABQ91695 | Abq91695 | M. capsul | c 229 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 157 | 16.2 | 60.0 | 729 | 7 | ACA30336 | Aca30336 | Prokaryote | c 230 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 158 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 231 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 159 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 232 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 160 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 233 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 161 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 234 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 162 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 235 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 163 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 236 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 164 | 16.2 | 60.0 | 866 | 2 | AAK79681 | Aak79681 | Cynomolgus | c 237 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 165 | 16.2 | 60.0 | 909 | 6 | ABZ11357 | Abz11357 | Human chromosome | c 238 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 166 | 16.2 | 60.0 | 1080 | 3 | AAK66376 | Aak66376 | Chitinase | c 239 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 167 | 16.2 | 60.0 | 1080 | 3 | AAK72652 | Aak72652 | Chitinase | c 240 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 168 | 16.2 | 60.0 | 1935 | 3 | ACA54037 | Aca54037 | Prokaryote | c 241 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |
| 169 | 16.2 | 60.0 | 2092 | 6 | ABL68555 | Abli68555 | Kidney calyculin | c 242 | 15.8 | 58.5 | 1632 | 3 | AAK11732 | Aak11732 | Human endotoxin |

243 15.8 58.5 3831 5 AAS71727 DNA encod
 244 15.8 58.5 3949 6 ABK89101 Aspergill
 245 15.8 58.5 4312 7 ABZ42855 Human met
 c 246 15.8 58.5 4519 7 ABT43574 Human cla
 c 247 15.8 58.5 4539 9 ADC17324 Human IFN
 c 248 15.8 58.5 5883 3 AAA88981 Toxoplas
 c 249 15.8 58.5 6560 4 ABL05125 Drosophil
 c 250 15.8 58.5 6774 4 ABL20811 Drosophil
 251 15.8 58.5 8411 5 ABA18086 Human ner
 252 15.8 58.5 12017 7 ABS58302 Novel hum
 253 15.8 58.5 12438 4 ABL20810 Drosophil
 254 15.8 58.5 16092 4 ABL05124 Drosophil
 255 15.8 58.5 21565 8 ADA02555 Mouse Ly6
 256 15.8 58.5 21565 9 ADB72293 Mouse Ly6
 257 15.8 58.5 24652 8 ADA02513 Mouse Ras
 258 15.8 58.5 24652 9 ADB72251 Mouse Ras
 c 259 15.8 58.5 34450 8 ACF04005 Human CRF
 c 260 15.8 58.5 34757 8 ACF04006 Human CRF
 c 261 15.8 58.5 38918 9 ADC87242 Human GPC
 c 262 15.8 58.5 89047 4 AAF28547 Genomic f
 c 263 15.8 58.5 90583 7 ACD13447 Human DNA
 264 15.8 58.5 110000 2 AAX91990_01
 c 265 15.8 58.5 110000 6 AQC67196 Listeria
 c 266 15.8 58.5 110000 6 ABQ69245_24
 c 267 15.8 58.5 110000 6 ABQ69245_25
 c 268 15.6 57.8 28 6 ABK66062 Human gen
 c 269 15.6 57.8 100 7 ACD76997 E. coli K
 c 270 15.6 57.8 256 6 ABL76576 Corn tass
 271 15.6 57.8 270 7 ABX82202 Corn ear-
 272 15.6 57.8 313 7 ABX85407 Corn ear-
 c 273 15.6 57.8 347 3 AAA30900 Breast ca
 c 274 15.6 57.8 360 6 ABV78204 Human ID3
 c 275 15.6 57.8 360 6 ABZ35780 Human ID3
 c 276 15.6 57.8 360 6 ABX10023 Human ID3
 c 277 15.6 57.8 360 6 ABL91745 Human pol
 c 278 15.6 57.8 429 8 ACH21548 Human adu
 c 279 15.6 57.8 458 4 AAL82306 Human pol
 c 280 15.6 57.8 465 8 ACH44238 Human foe
 c 281 15.6 57.8 495 4 AAI22113 Probe #12
 c 282 15.6 57.8 495 4 ABA67192 Human foe
 c 283 15.6 57.8 495 4 AAI47408 Probe #16
 c 284 15.6 57.8 495 4 ABA49278 Human bre
 c 285 15.6 57.8 495 4 ABA34287 Probe #12
 c 286 15.6 57.8 495 4 AKA41370 Human bon
 c 287 15.6 57.8 495 4 AKA15636 Human bra
 c 288 15.6 57.8 495 4 ABA40961 Human liv
 c 289 15.6 57.8 495 5 AAI07811 Probe #78
 c 290 15.6 57.8 495 6 ABS15375 Human gen
 c 291 15.6 57.8 687 3 AAF13257 Aspergill
 c 292 15.6 57.8 756 2 AQA20191 Ascorbic
 c 293 15.6 57.8 840 6 ABN67629 Streptoco
 c 294 15.6 57.8 843 7 ACA50767 Prokaryot
 c 295 15.6 57.8 982 2 AQA44245 HEIR-1 ge
 c 296 15.6 57.8 982 6 ABAQ88114 Human ost
 c 297 15.6 57.8 982 7 ACC46763 Human COP
 c 298 15.6 57.8 1020 7 ACC42715 Epoxide h
 c 299 15.6 57.8 1194 4 AAF88323 S. spinos
 c 300 15.6 57.8 1203 6 ABAQ88115 Human ost

ALIGNMENTS

RESULT 1
 AAI99682_25/c
 Continuation (26 of 45) of AAI99682 from base 2500001 (Mycobacterium tuberculosis strain
 WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
 WP Fragment Name Begin End
 WP AAI99682_00 1 110000
 WP AAI99682_01 100001 210000
 WP AAI99682_02 200001 310000
 WP AAI99682_03 300001 410000
 WP AAI99682_04 400001 510000
 WP AAI99682_05 500001 610000
 WP AAI99682_06 600001 710000
 WP AAI99682_07 700001 810000
 WP AAI99682_08 800001 910000
 WP AAI99682_09 900001 1010000
 WP AAI99682_10 1000001 1110000
 WP AAI99682_11 1100001 1210000
 WP AAI99682_12 1200001 1310000
 WP AAI99682_13 1300001 1410000
 WP AAI99682_14 1400001 1510000
 WP AAI99682_15 1500001 1610000
 WP AAI99682_16 1600001 1710000
 WP AAI99682_17 1700001 1810000

WP AAI99682_05 500001 610000
 WP AAI99682_06 600001 710000
 WP AAI99682_07 700001 810000
 WP AAI99682_08 800001 910000
 WP AAI99682_09 900001 1010000
 WP AAI99682_10 1000001 1110000
 WP AAI99682_11 1100001 1210000
 WP AAI99682_12 1200001 1310000
 WP AAI99682_13 1300001 1410000
 WP AAI99682_14 1400001 1510000
 WP AAI99682_15 1500001 1610000
 WP AAI99682_16 1600001 1710000
 WP AAI99682_17 1700001 1810000
 WP AAI99682_18 1800001 1910000
 WP AAI99682_19 1900001 2010000
 WP AAI99682_20 2000001 2110000
 WP AAI99682_21 2100001 2210000
 WP AAI99682_22 2200001 2310000
 WP AAI99682_23 2300001 2410000
 WP AAI99682_24 2400001 2510000
 WP AAI99682_25 2500001 2610000
 WP AAI99682_26 2600001 2710000
 WP AAI99682_27 2700001 2810000
 WP AAI99682_28 2800001 2910000
 WP AAI99682_29 2900001 3010000
 WP AAI99682_30 3000001 3110000
 WP AAI99682_31 3100001 3210000
 WP AAI99682_32 3200001 3310000
 WP AAI99682_33 3300001 3410000
 WP AAI99682_34 3400001 3510000
 WP AAI99682_35 3500001 3610000
 WP AAI99682_36 3600001 3710000
 WP AAI99682_37 3700001 3810000
 WP AAI99682_38 3800001 3910000
 WP AAI99682_39 3900001 4010000
 WP AAI99682_40 4000001 4110000
 WP AAI99682_41 4100001 4210000
 WP AAI99682_42 4200001 4310000
 WP AAI99682_43 4300001 4410000
 WP AAI99682_44 4400001 4411529

Query Match 100.0%; Score 27; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 0.0038;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CACCCGGCTTTAGGATCGACACCTGA 27
 |||||
 Db 104297 CACCCGGCTTTAGGATCGACACCTGA 104271

RESULT 2
 AAI99682_26/c
 Continuation (27 of 45) of AAI99682 from base 2600001 (Mycobacterium tuberculosis strain
 WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
 WP Fragment Name Begin End
 WP AAI99682_00 1 110000
 WP AAI99682_01 100001 210000
 WP AAI99682_02 200001 310000
 WP AAI99682_03 300001 410000
 WP AAI99682_04 400001 510000
 WP AAI99682_05 500001 610000
 WP AAI99682_06 600001 710000
 WP AAI99682_07 700001 810000
 WP AAI99682_08 800001 910000
 WP AAI99682_09 900001 1010000
 WP AAI99682_10 1000001 1110000
 WP AAI99682_11 1100001 1210000
 WP AAI99682_12 1200001 1310000
 WP AAI99682_13 1300001 1410000
 WP AAI99682_14 1400001 1510000
 WP AAI99682_15 1500001 1610000
 WP AAI99682_16 1600001 1710000
 WP AAI99682_17 1700001 1810000

WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Query Match 100.0%; Score 27; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27
DB 4297 CACCCGCGCTTTAGGATCGACACCTGA 4271

RESULT 3
AAI99683_25/c
Continuation (26 of 44) of AAI99683 from base 2500001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP Fragment Name Begin End
WP AAI99683_01 1000001 1100000
WP AAI99683_02 2000001 2100000
WP AAI99683_03 3000001 3100000
WP AAI99683_04 4000001 4100000
WP AAI99683_05 5000001 5100000
WP AAI99683_06 6000001 6100000
WP AAI99683_07 7000001 7100000
WP AAI99683_08 8000001 8100000
WP AAI99683_09 9000001 9100000
WP AAI99683_10 10000001 10100000
WP AAI99683_11 11000001 11100000
WP AAI99683_12 12000001 12100000
WP AAI99683_13 13000001 13100000
WP AAI99683_14 14000001 14100000
WP AAI99683_15 15000001 15100000
WP AAI99683_16 16000001 16100000
WP AAI99683_17 17000001 17100000
WP AAI99683_18 18000001 18100000
WP AAI99683_19 19000001 19100000
WP AAI99683_20 20000001 20100000
WP AAI99683_21 21000001 21100000
WP AAI99683_22 22000001 22100000
WP AAI99683_23 23000001 23100000
WP AAI99683_24 24000001 24100000
WP AAI99683_25 25000001 25100000
WP AAI99683_26 26000001 26100000
WP AAI99683_27 27000001 27100000
WP AAI99683_28 28000001 28100000
WP AAI99683_29 29000001 29100000
WP AAI99683_30 30000001 30100000

WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 100.0%; Score 27; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27
DB 100107 CACCCGCGCTTTAGGATCGACACCTGA 100081

RESULT 4
AAI99683_26/c
Continuation (27 of 44) of AAI99683 from base 2600001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP Fragment Name Begin End
WP AAI99683_00 1000001 1100000
WP AAI99683_01 2000001 2100000
WP AAI99683_02 3000001 3100000
WP AAI99683_03 4000001 4100000
WP AAI99683_04 5000001 5100000
WP AAI99683_05 6000001 6100000
WP AAI99683_06 7000001 7100000
WP AAI99683_07 8000001 8100000
WP AAI99683_08 9000001 9100000
WP AAI99683_09 10000001 10100000
WP AAI99683_10 11000001 11100000
WP AAI99683_11 12000001 12100000
WP AAI99683_12 13000001 13100000
WP AAI99683_13 14000001 14100000
WP AAI99683_14 15000001 15100000
WP AAI99683_15 16000001 16100000
WP AAI99683_16 17000001 17100000
WP AAI99683_17 18000001 18100000
WP AAI99683_18 19000001 19100000
WP AAI99683_19 20000001 20100000
WP AAI99683_20 21000001 21100000
WP AAI99683_21 22000001 22100000
WP AAI99683_22 23000001 23100000
WP AAI99683_23 24000001 24100000
WP AAI99683_24 25000001 25100000
WP AAI99683_25 26000001 26100000
WP AAI99683_26 27000001 27100000
WP AAI99683_27 28000001 28100000
WP AAI99683_28 29000001 29100000
WP AAI99683_29 30000001 30100000
WP AAI99683_30 31000001 31100000
WP AAI99683_31 32000001 32100000
WP AAI99683_32 33000001 33100000
WP AAI99683_33 34000001 34100000
WP AAI99683_34 35000001 35100000
WP AAI99683_35 36000001 36100000
WP AAI99683_36 37000001 37100000
WP AAI99683_37 38000001 38100000
WP AAI99683_38 39000001 39100000
WP AAI99683_39 40000001 40100000
WP AAI99683_40 41000001 41100000
WP AAI99683_41 42000001 42100000
WP AAI99683_42 43000001 43100000
WP AAI99683_43 44000001 4403765

Query Match 100.0%; Score 27; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTGA 27
DB 107 CACCGCGCTTTAGGATCGACACCTGA 81

RESULT 5
ID ABZ16792 standard; DNA; 2000 BP.
XX AC ABZ16792;
XX DT 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4597.
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS Arabidopsis thaliana.
XX PN WO200216655-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US026685.
XX PR 24-AUG-2000; 2000US-0227866P.
XX PR 26-JAN-2001; 2001US-0264647P.
XX PR 22-JUN-2001; 2001US-0300111P.
XX PA (SCRI) SCRIPPS RES INST.
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX PI Harper JF, Kreps J, Wang X, Zhu T;
XX WPI; 2002-304127/34.
XX PT Identifying a stress condition to which a plant cell has been exposed and
XX PT producing plants with increased tolerance to these abiotic stresses.
XX PS Claim 144; SEQ ID NO 4597; 577pp + Sequence Listing; English.
XX CC The invention relates to identifying a stress condition to which a plant
XX CC cell has been exposed, comprising: (a) contacting nucleic acid
XX CC representative of expressed polynucleotides in the plant cell with an
XX CC array or probes representative of the plant cell genome; and (b)
XX CC detecting a profile of expressed polynucleotides in the plant cell
XX CC characteristic of a stress response. The method is useful in the
XX CC production of transgenic plants, cells and seeds and in producing plants
XX CC with increased tolerance to abiotic stress. The present sequence is that
XX CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
XX CC in methods of the invention. Note: The sequence data for this patent is
XX CC not represented in the printed specification but is based on sequence
XX CC information supplied to Derwent by the European Patent Office
XX SQ Sequence 2000 BP; 634 A; 351 C; 308 G; 707 T; 0 U; 0 Other;

Query Match 68.9%; Score 18.6; DB 6; Length 2000;
Best Local Similarity 84.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTG 26
DB 783 ACCCGCGCTTTAGGATCGACACCTG 807

RESULT 6
ID AAH67093/c standard; DNA; 1287 BP.
XX

AAH67093;
26-SEP-2001 (first entry)
C glutamicum coding sequence fragment SEQ ID NO: 2128.
Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis; ds.
Corynebacterium glutamicum.
EPI109790-A2.
20-JUN-2001.
18-DEC-2000; 2000EP-00127688.
16-DEC-1999; 99JP-00377484.
07-APR-2000; 2000JP-00159162.
03-AUG-2000; 2000JP-00280988.
(KYOW) KYOWA HAKKO KOGYO KK.
Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
Tateishi N, Senoh A, Ikeda M, Ozaki A;
WPI; 2001-376931/40.
P-PSDE; AAG91874.
Novel polynucleotides derived from Coryneform bacteria, for identifying
mutation point of a gene, measuring expression of a gene, analyzing
expression profile or pattern of a gene and identifying homologous gene.
Claim 8; SEQ ID NO 2128; 246pp + Sequence Listing; English.
The present invention provides a number of nucleotide and protein
sequences from the Coryneform bacterium Corynebacterium glutamicum. These
are useful for identifying the mutation point of a gene derived from a
mutant of coryneform bacterium, measuring expression amount and analysing
the expression profile or expression pattern of a gene derived from
Coryneform bacterium, and identifying a homologue of a gene derived from
Coryneform bacterium. Coryneform bacteria are useful for producing amino
acids, nucleic acids, vitamins, saccharides and organic acids,
particularly L-lysine. The present sequence is a nucleic acid described
in the exemplification of the invention. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from the European Patent Office
Sequence 1287 BP; 258 A; 392 C; 338 G; 299 T; 0 U; 0 Other;

Query Match 66.7%; Score 18; DB 5; Length 1287;
Best Local Similarity 80.8%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTG 26
DB 90 CACCGCGCTTTAGGATCGACACCTG 65

RESULT 7
ID AAF68053/c standard; DNA; 1335 BP.
XX AC AAF68053;
XX DT 11-APR-2001 (first entry)
XX DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:621.
XX KW Corynebacterium glutamicum; brevivacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;

CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC Coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office

XX
SQ Sequence 349980 BP; 86896 A; 98023 C; 80939 G; 84122 T; 0 U; 0 Other;

Query Match 66.7%; Score 18; DB 5; Length 349980;
Best Local Similarity 80.8%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTG 26
||| ||||| ||||| |||||
Db 248517 CAGGCGCTCTTTAGGAGCGACAAC TG 248542

RESULT 10
AAS22935
ID ID AAS22935 standard; DNA; 766 BP.
XX AC AC AAS22935;
XX XX
DT 24-OCT-2001 (first entry)
XX XX
XX DNA encoding novel bone marrow polypeptide #29.
DE XX
XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID; ds.
XX XX
OS Homo sapiens.
OS
PN WC2001157187-A2.
XX XX
PD 09-AUG-2001.
XX XX
PF 05-FEB-2001; 2001WO-US0003782.
XX XX
PR 03-FEB-2000; 2000US-00496914.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 30-NOV-2000; 2000US-0250683P.
XX XX
PA (HYSE-) HYSEQ INC.
XX XX
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
PI
XX
XX WPI; 2001-488875/53.
DR P-PSDB; AAU14630.
DR
XX
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
PT gene therapy.
PT
PS
PS Claim 1; Page 171; 392pp; English.
XX
XX AAS22907-AAS23099 represent nucleic acids encoding novel bone marrow
CC polypeptides. The nucleic acids and corresponding proteins may be used in
CC the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations or
CC deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein. The
CC nucleic acid and its complementary sequences may also be used as DNA

DR WPI; 2001-071485/08.
 DR P-PSDE; AAB76819.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 DR and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation.
 XX
 XX Claim 3; Page 1036-1038; 1119pp; English.
 XX
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention
 XX
 XX Sequence 1410 BP; 290 A; 427 C; 364 G; 329 T; 0 U; 0 Other;
 SQ
 Query Match 66.7%; Score 18; DB 4; Length 1410;
 Best Local Similarity 80.8%; Pred. No. 69;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 CACCGCGCTTTAGGATCGACACTG 26
 |||||
 Db 190 CAGCGCTCTTTAGGAGCGCACTG 165
 |||||
 RESULT 9
 AAH68530
 ID AAH68530 standard; DNA; 349980 BP.
 XX
 AC AAH68530;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 7065.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-00127688.
 XX
 PR 16-DEC-1999; 99JP-00377484.
 PR 07-APR-2000; 2000JP-00159162.
 PR 03-AUG-2000; 2000JP-00280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analyzing
 PT expression profile or pattern of a gene and identifying homologous gene.
 XX
 PS Disclosure; SEQ ID NO 7065; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a

patient's own production of the polypeptide. Additionally, the nucleic acids may be used to produce the polypeptides, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The nucleic acid and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and therefore which patients may be in need of restorative therapy. The proteins may also be used as antigens in the production of antibodies against bone marrow proteins and in assays to identify modulators of their expression and activity. The anti-bone marrow protein antibodies and antagonists may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the protein in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins may be used to regulate haematopoiesis activity, and consequently in the treatment of myeloid or lymph cell disorders; in tissue regeneration, such as wound healing; as a nutritional supplement; and in treatment of immune disorders such as severe combined immunodeficiency (SCID)

```
Query Match          65.2%; Score 17.6; DB 4; Length 771;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

109 CACCCCGCCTCAGGATCGAGACC 86

ABS56529;
27-JAN-2003 (first entry)

27-JAN-2003 (first entry)

us; genes; phenylalanine
PAH; para-hydroxycinnamic
cinnamic acid; PHCA; liquid
Chromobacterium violaceum.

| Key | Location/Qualifiers |
|-----|---------------------|
| chs | 1 891 |

```

key
CDS
location/qualifiers
1..891
/*tag= a
/product= "phenylalanine ammonia-lyase"

```

WO200290523-A2.

WO200290523-A2.
14-NOV-2002.

14-NOV-2002.

03-MAY-2002:

04-MAY-2001: 3001115-0389701P

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

100

WPI; 2003-058928/05.
P-PSDB; ABG71947.

F-300B, 400, 1000, 1000.

Example 1; Page 45; 69pp; English.

CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
 CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
 CC providing the recombinant host; and (2) growing the recombinant organism
 CC in the presence of a fermentable carbon substrate and thus producing para
 CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
 CC providing a recombinant organism comprising at least one gene encoding a
 CC phenylalanine hydroxylase activity; and (2) growing the recombinant
 CC organism in the presence of a fermentable carbon substrate to produce
 CC tyrosine. The recombinant host cell is used for producing para
 CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
 CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
 CC monomer for the production of liquid crystal polymer. The present
 CC sequence is C. violaceum DNA encoding Phenylalanine ammonia-lyase, PAH,
 CC an enzyme which catalyses the conversion of phenylalanine into cinnamic
 CC acid
 CC
 XX
 SQ Sequence 894 BP; 167 A; 301 C; 274 G; 152 T; 0 U; 0 Other;
 Query Match 65.2%; Score 17.6; DB 7; Length 894;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 ACCCGCGCTTTAGGATCGACACCT 25
 DB 686 ACACGCGCTACCGGATCGACACCT 709
 RESULT 13
 ID ACA25099/c
 AC ACA25099 standard; DNA; 933 BP.
 XX
 AC ACA25099;
 XX
 DT 19-JUN-2003 (first entry)
 DE Prokaryotic essential gene #6756.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Burkholderia fungorum.
 XX
 FN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU21229.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 12969; 1766pp; English.
 CC
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 933 BP; 164 A; 293 C; 321 G; 155 T; 0 U; 0 Other;
 Query Match 65.2%; Score 17.6; DB 7; Length 933;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 3 CCGCGCGCTTTAGGATCGACACCTG 26
 DB 486 CTCGAGCTTCAGGATCGTCACTG 463
 RESULT 14
 ID AAS73721
 AC AAS73721 standard; cDNA; 1195 BP.
 XX
 AC AAS73721;
 XX
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #9525.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 FN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG09534.
 XX
 DR
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 1; SEQ ID NO 9525; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (II) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic

CC coding sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

XX SQ Sequence 1195 BP; 208 A; 373 C; 372 G; 241 T; 0 U; 1 Other;

Query Match 65.2%; Score 17.6; DB 5; Length 1195;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24

Db 769 CCCCCGCGCTTTATGACCGACCCC 792

RESULT 15

ABNS9815

ID ABNS9815 standard; cDNA; 4919 BP.

XX

XX AC ABNS9815;

XX

XX DT 28-JUN-2002 (first entry)

XX

XX DE Novel human coding sequence SEQ ID NO: 226.

XX

XX KW Human; antianaemic; vulnary; antiinflammatory; immunomodulator;

XX KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;

XX KW neuroprotective; antiparkinsonian; protein therapy; EST;

XX KW expressed sequence tag; gene; ss.

XX

XX OS Homo sapiens.

XX

XX PN WO200222660-A2.

XX

XX PD 21-MAR-2002.

XX

XX PF 10-SEP-2001; 2001WO-US026015.

XX

XX PR 11-SEP-2000; 2000US-00659671.

XX

XX PA (HYSE-) HYSEQ INC.

XX

XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QX, Ren F;

XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX

XX DR WPI; 2002-292408/33.

XX DR P-PSDB; ABB97402.

XX

XX PT An isolated polynucleotide for treating diseases associated with its

XX encoded polypeptide such as cancer and multiple sclerosis.

XX

XX PS Claim 1; SEQ ID NO 226; 509pp; English.

XX

CC The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke

CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.

CC rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a coding sequence of the

CC invention

XX SQ Sequence 4919 BP; 1366 A; 1264 C; 1177 G; 1112 T; 0 U; 0 Other;

Query Match 65.2%; Score 17.6; DB 6; Length 4919;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24

Db 1446 CACCCGCGCTTCAGATCGAGACC 1469

RESULT 16

ACA38547/C

ID ACA38547 standard; DNA; 1789 BP.

XX

XX AC ACA38547;

XX

XX DT 19-JUN-2003 (first entry)

XX

XX DE Prokaryotic essential gene #20204.

XX

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX

XX OS Mycobacterium bovis.

XX

XX PN WO200277183-A2.

XX

XX PD 03-OCT-2002.

XX

XX PF 21-MAR-2002; 2002WO-US009107.

XX

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342823P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX

XX PA (ELIT-) ELITRA PHARM INC.

XX

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX

XX DR WPI; 2003-029926/02.

XX DR P-PSDB; ABU34677.

XX

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX

XX PS Claim 14; SEQ ID NO 26417; 1766pp; English.

XX

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1789 BP; 362 A; 573 C; 586 G; 268 T; 0 U; 0 Other;

Query Match 64.4%; Score 17.4; DB 7; Length 1789;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACTGA 27
Db 1293 CGCCGCGCGCAGGATCGTCACTTA 1267

RESULT 17

AAH52081/c
ID AAH52081 standard; DNA; 2283 BP.

XX AAH52081;

DT 04-SEP-2001 (first entry)

DE Mycobacterium tuberculosis potential drug target gene SEQ ID 135.

XX Drug target; growth; organism viability; characterisation; ds.

XX Mycobacterium tuberculosis.

XX WO200135317-A1.

XX 17-MAY-2001.

XX 13-NOV-2000; 2000WO-US031152.

XX 12-NOV-1999; 99US-0165086P.

XX 12-NOV-1999; 99US-0165124P.

XX 01-FEB-2000; 2000US-0179531P.

XX (REGC) UNIV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI; 2001-329193/34.

XX P-PSDB; AAG81230.

XX Identifying nucleotide or polypeptide sequence for use as drug target,

XX involves providing algorithm that analyzes a functional relationship

XX between nucleotide or polypeptide sequences, and comparing the sequences.

XX Disclosure; Page 143-144; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or

CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism

Sequence 2283 BP; 452 A; 757 C; 745 G; 329 T; 0 U; 0 Other;

Query Match 64.4%; Score 17.4; DB 4; Length 2283;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACTGA 27

Db 1355 CGCCGCGCGCAGGATCGTCACTTA 1329

RESULT 18

ACA40856/c

ID ACA40856 standard; DNA; 2283 BP.

XX ACA40856;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #22513.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Mycobacterium tuberculosis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-0299926/02.

XX P-PSDB; ABU36986.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 28726; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC on a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. Note: The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2283 BP; 452 A; 757 C; 745 G; 329 T; 0 U; 0 Other;
SQ Query Match 64.4%; Score 17.4; DB 7; Length 2283;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
Db 1355 CGCCGCGCGCCAGGATCGTACCTTA 1329

RESULT 19
ABL05719/c
ID ABL05719 standard; cDNA; 2937 BP.
XX ABL05719;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 11639.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB61616.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 11639; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2937 BP; 643 A; 808 C; 781 G; 705 T; 0 U; 0 Other;
SQ Query Match 64.4%; Score 17.4; DB 4; Length 2937;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCG 19
Db 1491 CACCCGCTTTAGGATCG 1473

RESULT 20
ABL05718
ID ABL05718 standard; cDNA; 5484 BP.
XX ABL05718;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 11636.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB61615.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 11636; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 5484 BP; 1530 A; 1265 C; 1329 G; 1360 T; 0 U; 0 Other;
SQ Query Match 64.4%; Score 17.4; DB 4; Length 5484;
Best Local Similarity 94.7%; Pred. No. 1.7e+02;

| Query Match | Best Local Similarity | Score 17.4; DB 4; Length 110000; |
|--------------------------|------------------------------------|----------------------------------|
| Matches 21; Conservative | 77.8%; Pred. No. 2.5e+02; | |
| | 0; Mismatches 6; Indels 0; Gaps 0; | |
| AAI99683_01 | 100001 | 210000 |
| AAI99683_02 | 200001 | 310000 |
| AAI99683_03 | 300001 | 410000 |
| AAI99683_04 | 400001 | 510000 |
| AAI99683_05 | 500001 | 610000 |
| AAI99683_06 | 600001 | 710000 |
| AAI99683_07 | 700001 | 810000 |
| AAI99683_08 | 800001 | 910000 |
| AAI99683_09 | 900001 | 1010000 |
| AAI99683_10 | 1000001 | 1110000 |
| AAI99683_11 | 1100001 | 1210000 |
| AAI99683_12 | 1200001 | 1310000 |
| AAI99683_13 | 1300001 | 1410000 |
| AAI99683_14 | 1400001 | 1510000 |
| AAI99683_15 | 1500001 | 1610000 |
| AAI99683_16 | 1600001 | 1710000 |
| AAI99683_17 | 1700001 | 1810000 |
| AAI99683_18 | 1800001 | 1910000 |
| AAI99683_19 | 1900001 | 2010000 |
| AAI99683_20 | 2000001 | 2110000 |
| AAI99683_21 | 2100001 | 2210000 |
| AAI99683_22 | 2200001 | 2310000 |
| AAI99683_23 | 2300001 | 2410000 |
| AAI99683_24 | 2400001 | 2510000 |
| AAI99683_25 | 2500001 | 2610000 |
| AAI99683_26 | 2600001 | 2710000 |
| AAI99683_27 | 2700001 | 2810000 |
| AAI99683_28 | 2800001 | 2910000 |
| AAI99683_29 | 2900001 | 3010000 |
| AAI99683_30 | 3000001 | 3110000 |
| AAI99683_31 | 3100001 | 3210000 |
| AAI99683_32 | 3200001 | 3310000 |
| AAI99683_33 | 3300001 | 3410000 |
| AAI99683_34 | 3400001 | 3510000 |
| AAI99683_35 | 3500001 | 3610000 |
| AAI99683_36 | 3600001 | 3710000 |
| AAI99683_37 | 3700001 | 3810000 |
| AAI99683_38 | 3800001 | 3910000 |
| AAI99683_39 | 3900001 | 4010000 |
| AAI99683_40 | 4000001 | 4110000 |
| AAI99683_41 | 4100001 | 4210000 |
| AAI99683_42 | 4200001 | 4310000 |
| AAI99683_43 | 4300001 | 4403765 |

Query Match 64.4%; Score 17.4; DB 4; Length 110000;

Best Local Similarity 77.8%; Pred. No. 2.5e+02;

Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 CACCCGCGCTTAGGATCGACACCTGA 27

Db 49594 CGCCGCGCGCCAGGATCGTCACCTTA 43620

RESULT 23

ABK74646/C

ID ABK74646 standard; DNA; 999 BP.

XX AC

XX ABK74646;

XX

DT 13-AUG-2002 (first entry)

XX

DE Bacillus licheniformis genomic sequence tag (GSR) #1937.

XX

KW Differential gene expression; genomic sequenced tag; GSR;

KW altered culture condition; environmental stress;

KW physiological provocation; ds.

XX

OS Bacillus licheniformis.

XX

FN WO200229113-A2.

XX

PD 11-APR-2002.

XX

PF 05-OCT-2003; 2001WO-US031437.
XX
PR 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
PI Berka R, Clausen IG;
XX WPI; 2002-416684/44.
XX
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX
PS Claim 4; SEQ ID NO 1937; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GSTR), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions.
CC environmental stress or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GSTR) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 999 BP; 298 A; 228 C; 251 G; 222 T; 0 U; 0 Other;

Query Match 63.7%; Score 17.2; DB 6; Length 999;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACA 22
DB 960 CACACGCGCTTTAGGATCCATA 939

RESULT 24
ACA36049
ID ACA36049 standard; DNA; 1257 BP.
XX
AC ACA36049;
XX
XX 19-JUN-2003 (first entry)
DT
XX Prokaryotic essential gene #17706.
DE
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Klebsiella pneumoniae.
XX WO20027183-A2.
XX
XX 03-OCT-2002.
PD
XX 21-MAR-2002; 2002WO-US009107.
XX
XX

21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR P-PSDB; ABU321179.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 23919; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1257 BP; 201 A; 403 C; 422 G; 231 T; 0 U; 0 Other;

Query Match 63.7%; Score 17.2; DB 7; Length 1257;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CCGCGCTTTAGGATCGACACCT 25
DB 340 CCGCGCTTTAGGATCGACACCT 361

RESULT 25
AAK44543/C
ID AAK44543 standard; DNA; 279 BP.
XX
AC AAK44543;
XX
XX 06-NOV-2001 (first entry)
DT
XX Human bone marrow expressed single exon probe SEQ ID NO: 19100.
DE
XX

KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 03-AUG-2000; 2000US-0234687P.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
XX
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 19100; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
XX Sequence 279 BP; 76 A; 64 C; 67 G; 72 T; 0 U; 0 Other;
SQ
Query Match 63.0%; Score 17; DB 4; Length 279;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCCGCGCTTAGGATCGACACCTG 26
DB 36 ACCCTCTCTTTGGGATCCCACTG 12
RESULT 26
ABS44201/c
ID ABS44201 standard; DNA; 279 BP.
XX
AC ABS44201;
XX
XX 25-FEB-2003 (first entry)
DT
DE Human liver single exon probe, SEQ ID No 19191.
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
XX Homo sapiens.
XX
XX WO200157273-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR

PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488998/53.
XX
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 4; SEQ ID NO 19191; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 279 BP; 76 A; 64 C; 67 G; 72 T; 0 U; 0 Other;
SQ
Query Match 63.0%; Score 17; DB 4; Length 279;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCCGCGCTTAGGATCGACACCTG 26
DB 36 ACCCTCTCTTTGGGATCCCACTG 12
RESULT 27
ABS18780/c
ID ABS18780 standard; DNA; 279 BP.
XX
AC ABS18780;
XX
XX 19-AUG-2002 (first entry)
DT
DE Human genome-derived single exon probe ORF from lung SEQ ID No 18771.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioma; lymphoma;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
XX Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX

PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.
 PT
 PS Claim 4; SEQ ID NO 18771; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC ; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA; and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a single exon probe open reading frame of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 279 BP; 76 A; 64 C; 67 G; 72 T; 0 U; 0 Other;
 Query Match 63.0%; Score 17; DB 6; Length 279;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 2 ACCCGCGCTTTAGGATCGACACCTG 26
 DB 36 ACCTCTCTTGGATCCACACTG 12
 RESULT 28
 ABS69598
 ID ABS69598 standard; DNA; 349 BP.
 XX
 AC ABS69598;
 XX
 DT 21-NOV-2002 (first entry)

XX
 DE Novel murine polynucleotide isolated using gene trap technology #661.
 XX
 KW Mouse; gene trapped sequence; GTS; functional genomic analysis;
 KW phase display system; gene chip; temporal gene expression;
 KW tissue specific gene expression; antisense inhibition; gene targeting;
 KW development disorder; cell differentiation disorder; aging; cancer;
 KW autoimmune disease; lupus; inflammatory disorder; skin disorder;
 KW degenerative disorder; ds.
 XX
 OS Mus musculus.
 XX
 PN US2002102543-A1.
 XX
 PD 01-AUG-2002.
 XX
 PF 30-NOV-2000; 2000US-00728445.
 XX
 PR 01-DEC-1999; 99US-0168358P.
 XX
 PA (FRIE/) FRIEDRICH G.
 PA (ZAMB/) ZAMBROWICZ B.
 PA (SAND/) SANDS A T.
 PI Friedrich G, Zambrowicz B, Sands AT;
 DR WPI; 2002-690598/74.
 XX
 PT Novel murine polynucleotides that individually identify novel genes into
 PT which a retroviral gene trap vector has integrated, useful in genomic
 PT analysis and in discovery, development of therapeutic and diagnostic
 PT agents.
 XX
 PS Claim 1; Page 222; 295pp; English.
 XX
 CC The invention describes an isolated murine polynucleotide (I) comprising
 CC a contiguous stretch of at least 60 nucleotides of one of 265-677
 CC nucleotide 991 OMNI-BANK gene trapped sequences (GTS) (S), given in the
 CC specification. The novel genes and cells are useful in functional genomic
 CC analysis and in the discovery and development of new therapeutic and
 CC diagnostic agents and methods. (I) is useful for identifying the coding
 CC regions of the murine genome, to isolate cDNAs, genomic clones, or full-
 CC length genes/polynucleotides or homologues, heterologues, paralogues, or
 CC orthologues that are capable of hybridising to one or more of the GTSs
 CC under stringent conditions. (I) can be incorporated into a phage display
 CC system that can be used to screen for proteins, or other ligands, that
 CC are capable of binding an amino acid sequence encoded by an
 CC oligonucleotide or polynucleotide sequence in at least one of the TS
 CC sequences. (I) is useful in addressable arrays, such as gene chips, to
 CC identify and characterise temporal and tissue specific gene expression,
 CC to identify the gene of interest from many sources and for genetic
 CC manipulations such as antisense inhibition and gene targeting. Decreasing
 CC the level of expression of (I) and/or down regulating the activity of
 CC peptides or proteins encoded by (I) is useful for treating development
 CC and cell differentiation disorders, aging, cancer, autoimmune disease,
 CC lupus, inflammatory disorders, skin disorders and degenerative disorders.
 CC This sequence represents a murine cDNA isolated using gene trap
 CC technology
 XX
 SQ Sequence 349 BP; 76 A; 95 C; 97 G; 66 T; 0 U; 15 Other;
 Query Match 63.0%; Score 17; DB 6; Length 349;
 Best Local Similarity 80.0%; Pred. No. 1.8e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 1 CACCGCGCTTTAGGATCGACACCT 25
 DB 158 CACCGCGCTTCTCAGATCGACATCT 182
 RESULT 29
 RAK31460/c
 ID RAK31460 standard; DNA; 463 BP.

```
XX AAK31460;
AC
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 6017.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 6017; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 463 BP; 108 A; 121 C; 111 G; 123 T; 0 U; 0 Other;
Query Match 63.0%; Score 17; DB 4; Length 463;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCCGCGCTTTAGGATCGACACTG 26
DB 262 ACCCTCTCTTTGGGATCCCAACTG 238
RESULT 30
ABS31142/c
ID ABS31142 standard; DNA; 463 BP.
XX
AC ABS31142;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 6132.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
```

```
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488989/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 1; SEQ ID NO 6132; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences
XX
SQ Sequence 463 BP; 108 A; 121 C; 111 G; 123 T; 0 U; 0 Other;
Query Match 63.0%; Score 17; DB 4; Length 463;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACCCGCGCTTTAGGATCGACACTG 26
DB 262 ACCCTCTCTTTGGGATCCCAACTG 238
Search completed: June 20, 2004, 10:16:44
Job time : 177.268 secs
```



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1261.71 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-12
Perfect score: 27
Sequence: 1 caccgcgcttagatcacactga 27

Scoring table: IDENTITY_NUC
Gapcp 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

EST:

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estm:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_est3:*
- 12: gb_est4:*
- 13: gb_est5:*
- 14: gb_est6:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pla:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| 1 | 20.2 | 74.8 | 324 | 10 | BB504590 |
| C 2 | 19.6 | 72.5 | 630 | 13 | CA055313 |
| C 3 | 19.4 | 71.9 | 634 | 28 | AQ259762 |
| 4 | 19.2 | 71.1 | 1130 | 13 | BQ683442 |

| | | | | | | | | | | | | | |
|-----|------|------|------|----|-----------|-------------|-----|----|------|-----|----|----------|--------------------|
| 78 | 17.4 | 64.4 | 557 | 14 | CA755331 | BR0300180 | 151 | 17 | 63.0 | 244 | 11 | CNS08PML | BX009673 Single re |
| 79 | 17.4 | 64.4 | 565 | 28 | AQ355484 | CITBI-EL- | 152 | 17 | 63.0 | 249 | 13 | BQ979128 | BQ979128 OH17E12.Y |
| 80 | 17.4 | 64.4 | 573 | 28 | AZ756601 | ev13G10.Y | 153 | 17 | 63.0 | 256 | 9 | AA219969 | AA219969 MV63H04.Y |
| 81 | 17.4 | 64.4 | 578 | 14 | CB936801 | IPCGJX13 | 154 | 17 | 63.0 | 274 | 10 | BB567413 | BB567413 BB567413 |
| 82 | 17.4 | 64.4 | 516 | 28 | BZ103006 | BZ230-147 | 155 | 17 | 63.0 | 298 | 14 | W14809 | W14809 MD32G04.Y |
| 83 | 17.4 | 64.4 | 635 | 28 | BH281652 | BH230-128 | 156 | 17 | 63.0 | 338 | 14 | W12931 | W12931 MD32G04.Y |
| 84 | 17.4 | 64.4 | 636 | 29 | CE718914 | tigr-gss- | 157 | 17 | 63.0 | 345 | 10 | B2333640 | B2333640 B2333640 |
| 85 | 17.4 | 64.4 | 654 | 14 | CA294412 | SCGLV101 | 158 | 17 | 63.0 | 349 | 29 | CG532435 | CG532435 OST116301 |
| 86 | 17.4 | 64.4 | 675 | 28 | BH403540 | AG-ND-136 | 159 | 17 | 63.0 | 352 | 13 | BY339623 | BY339623 BY339623 |
| 87 | 17.4 | 64.4 | 685 | 29 | CE449767 | tigr-gss- | 160 | 17 | 63.0 | 354 | 9 | AU081974 | AU081974 AU081974 |
| 88 | 17.4 | 64.4 | 686 | 29 | BH372901 | AG-ND-136 | 161 | 17 | 63.0 | 385 | 9 | AI385593 | AI385593 ml53f01.Y |
| 89 | 17.4 | 64.4 | 690 | 13 | CE120332 | tigr-gss- | 162 | 17 | 63.0 | 397 | 29 | CG50685 | CG50685 OST157121 |
| 90 | 17.4 | 64.4 | 690 | 13 | BH548010 | GM880012B | 163 | 17 | 63.0 | 398 | 9 | AA041630 | AA041630 mJ03f08.Y |
| 91 | 17.4 | 64.4 | 715 | 10 | BH824053 | BH824053 | 164 | 17 | 63.0 | 401 | 13 | BY059141 | BY059141 BY059141 |
| 92 | 17.4 | 64.4 | 835 | 14 | CF355494 | V-B-128G1 | 165 | 17 | 63.0 | 404 | 13 | BY313275 | BY313275 BY313275 |
| 93 | 17.4 | 64.4 | 1066 | 12 | BG664932 | DRABRF10 | 166 | 17 | 63.0 | 421 | 10 | B739467 | B739467 B739467 |
| 94 | 17.4 | 64.4 | 1077 | 29 | CE811558 | ZMMBCC047 | 167 | 17 | 63.0 | 427 | 14 | W65993 | W65993 me09s03.Y |
| 95 | 17.2 | 63.7 | 166 | 10 | AH870591 | pa61G12.Y | 168 | 17 | 63.0 | 430 | 13 | BY081166 | BY081166 BY081166 |
| 96 | 17.2 | 63.7 | 187 | 10 | BZ243785 | DA85H09.Y | 169 | 17 | 63.0 | 434 | 14 | W98485 | W98485 mg20f11.Y |
| 97 | 17.2 | 63.7 | 200 | 13 | C28565 | C28565 Rice | 170 | 17 | 63.0 | 435 | 12 | B1220343 | B1220343 602935636 |
| 98 | 17.2 | 63.7 | 283 | 12 | BG232344 | pa68d10.Y | 171 | 17 | 63.0 | 437 | 10 | B848348 | B848348 B848348 |
| 99 | 17.2 | 63.7 | 307 | 10 | AW627142 | pa18d01.Y | 172 | 17 | 63.0 | 444 | 10 | B862117 | B862117 B862117 |
| 100 | 17.2 | 63.7 | 311 | 13 | BH038432 | DH09C09.H | 173 | 17 | 63.0 | 450 | 11 | CNS08KMX | AL389084 MCBC2F08 |
| 101 | 17.2 | 63.7 | 316 | 10 | AW700497 | pa36h04.Y | 174 | 17 | 63.0 | 456 | 14 | W13859 | W13859 mc36c01.Y |
| 102 | 17.2 | 63.7 | 316 | 10 | AW700848 | pa44h05.Y | 175 | 17 | 63.0 | 457 | 10 | B8333373 | B8333373 B8333373 |
| 103 | 17.2 | 63.7 | 316 | 10 | AW734994 | pa46b01.Y | 176 | 17 | 63.0 | 460 | 10 | B8333373 | B8333373 B8333373 |
| 104 | 17.2 | 63.7 | 322 | 10 | AW734998 | pa46b06.Y | 177 | 17 | 63.0 | 465 | 9 | AA221641 | AA221641 mv50C07.Y |
| 105 | 17.2 | 63.7 | 330 | 10 | BZ250381 | pa98e01.Y | 178 | 17 | 63.0 | 465 | 14 | W45960 | W45960 mc82a01.Y |
| 106 | 17.2 | 63.7 | 336 | 28 | AZ143316 | SP_0040.B | 179 | 17 | 63.0 | 466 | 12 | B245335 | B245335 949031C03 |
| 107 | 17.2 | 63.7 | 353 | 10 | BZ352555 | pa34g12.Y | 180 | 17 | 63.0 | 472 | 10 | AW538802 | AW538802 C0112A01- |
| 108 | 17.2 | 63.7 | 364 | 12 | BH077845 | pb20h02.Y | 181 | 17 | 63.0 | 477 | 11 | CNS09BTK | AL389084 MCBC2F08 |
| 109 | 17.2 | 63.7 | 376 | 10 | AW735241 | pa49e12.Y | 182 | 17 | 63.0 | 487 | 9 | AL389084 | AL389084 MCBC2F08 |
| 110 | 17.2 | 63.7 | 394 | 10 | BZ250028 | pa49b08.Y | 183 | 17 | 63.0 | 487 | 10 | BE647149 | BE647149 UI-N-BH1- |
| 111 | 17.2 | 63.7 | 397 | 13 | BQ666623 | pa49g10.Y | 184 | 17 | 63.0 | 490 | 10 | AW773035 | AW773035 925001G04 |
| 112 | 17.2 | 63.7 | 403 | 10 | BZ352430 | pa33c04.Y | 185 | 17 | 63.0 | 493 | 9 | AL927921 | AL927921 AL927921 |
| 113 | 17.2 | 63.7 | 408 | 10 | BH101391 | UI-R-BJ1- | 186 | 17 | 63.0 | 493 | 14 | CA565603 | CA565603 K0340B07- |
| 114 | 17.2 | 63.7 | 411 | 10 | AW627032 | pa23e04.Y | 187 | 17 | 63.0 | 496 | 28 | AQ999515 | AQ999515 RPI1-23-3 |
| 115 | 17.2 | 63.7 | 416 | 10 | AW627032 | pa23e04.Y | 188 | 17 | 63.0 | 499 | 9 | AA022332 | AA022332 mh85404.Y |
| 116 | 17.2 | 63.7 | 416 | 10 | BZ250494 | pa99g05.Y | 189 | 17 | 63.0 | 510 | 14 | R64936 | R64936 13440 Lamdd |
| 117 | 17.2 | 63.7 | 423 | 10 | BZ250749 | pa04a05.Y | 190 | 17 | 63.0 | 514 | 9 | AA237507 | AA237507 mx10e10.Y |
| 118 | 17.2 | 63.7 | 427 | 10 | AW588592 | pa15h05.Y | 191 | 17 | 63.0 | 514 | 10 | AW701380 | AW701380 uq88B05.X |
| 119 | 17.2 | 63.7 | 428 | 10 | BZ250161 | pa89f05.Y | 192 | 17 | 63.0 | 519 | 10 | B766415 | B766415 BB766415 |
| 120 | 17.2 | 63.7 | 448 | 12 | BH386369 | UI-R-CN1- | 193 | 17 | 63.0 | 520 | 29 | CE380046 | CE380046 tigr-gss- |
| 121 | 17.2 | 63.7 | 449 | 12 | BG376436 | UI-R-CU0- | 194 | 17 | 63.0 | 521 | 11 | CNS09CL3 | BX052387 Single re |
| 122 | 17.2 | 63.7 | 456 | 10 | AW627329 | pa28g10.Y | 195 | 17 | 63.0 | 524 | 11 | CNS09CL3 | BX042541 Single re |
| 123 | 17.2 | 63.7 | 456 | 10 | BZ249840 | pa89g09.Y | 196 | 17 | 63.0 | 527 | 29 | CG622383 | CG622383 OST121875 |
| 124 | 17.2 | 63.7 | 459 | 10 | BZ250798 | pa04a08.Y | 197 | 17 | 63.0 | 533 | 9 | AI552256 | AI552256 VF73404.Y |
| 125 | 17.2 | 63.7 | 467 | 10 | BZ249819 | pa86e05.Y | 198 | 17 | 63.0 | 537 | 14 | CA182127 | CA182127 SCBGT310 |
| 126 | 17.2 | 63.7 | 470 | 10 | AW870543 | pa16a02.Y | 199 | 17 | 63.0 | 537 | 9 | AA637516 | AA637516 vu10a04.Y |
| 127 | 17.2 | 63.7 | 484 | 10 | BZ250621 | pa02d10.Y | 200 | 17 | 63.0 | 538 | 13 | BX520418 | BX520418 BX520418 |
| 128 | 17.2 | 63.7 | 489 | 10 | AW700407 | pa1a12.Y | 201 | 17 | 63.0 | 551 | 13 | AA81181 | AA81181 v207c12.Y |
| 129 | 17.2 | 63.7 | 500 | 10 | AW588363 | pa12c04.Y | 202 | 17 | 63.0 | 552 | 10 | AW519606 | AW519606 UP36d10.Y |
| 130 | 17.2 | 63.7 | 502 | 12 | BH077610 | pa11f11.Y | 203 | 17 | 63.0 | 552 | 12 | BI679671 | BI679671 949031C03 |
| 131 | 17.2 | 63.7 | 510 | 12 | BH181947 | pa181947 | 204 | 17 | 63.0 | 563 | 14 | CK333832 | CK333832 H8262B08- |
| 132 | 17.2 | 63.7 | 516 | 10 | BH1818423 | UI-R-BJ1- | 205 | 17 | 63.0 | 572 | 10 | AW540159 | AW540159 C0129C03- |
| 133 | 17.2 | 63.7 | 524 | 12 | BH077440 | pa13c12.Y | 206 | 17 | 63.0 | 575 | 10 | BG077099 | BG077099 H3010G01- |
| 134 | 17.2 | 63.7 | 524 | 12 | BH077440 | pa13c12.Y | 207 | 17 | 63.0 | 575 | 11 | CG313887 | CG313887 tigr-gss- |
| 135 | 17.2 | 63.7 | 535 | 12 | BH077892 | pa21e10.Y | 208 | 17 | 63.0 | 580 | 29 | CE313887 | CE313887 tigr-gss- |
| 136 | 17.2 | 63.7 | 535 | 12 | BH077892 | pa21e10.Y | 209 | 17 | 63.0 | 581 | 9 | AA473043 | AA473043 v43308.Y |
| 137 | 17.2 | 63.7 | 540 | 12 | BH077845 | pa12d08.Y | 210 | 17 | 63.0 | 581 | 28 | AQ401049 | AQ401049 HS-5058.A |
| 138 | 17.2 | 63.7 | 541 | 12 | BH077845 | pa18g12.Y | 211 | 17 | 63.0 | 583 | 12 | BG277813 | BG277813 ux61f05.Y |
| 139 | 17.2 | 63.7 | 541 | 12 | BH077827 | pa20f05.Y | 212 | 17 | 63.0 | 586 | 14 | CB082687 | CB082687 hn55h06.Y |
| 140 | 17.2 | 63.7 | 552 | 12 | BH077773 | pa39h02.Y | 213 | 17 | 63.0 | 592 | 12 | BI797918 | BI797918 H091D09.E |
| 141 | 17.2 | 63.7 | 556 | 28 | BZ368110 | id1d10.9 | 214 | 17 | 63.0 | 594 | 10 | BE439656 | BE439656 HTML-323F |
| 142 | 17.2 | 63.7 | 573 | 9 | AA849040 | EST191802 | 215 | 17 | 63.0 | 596 | 13 | BQ206410 | BQ206410 UI-R-DZ1- |
| 143 | 17.2 | 63.7 | 667 | 9 | AI406406 | EST234692 | 216 | 17 | 63.0 | 600 | 28 | BH75241 | BH75241 AG-ND-109 |
| 144 | 17.2 | 63.7 | 684 | 10 | BE113133 | UI-R-BJ1- | 217 | 17 | 63.0 | 605 | 9 | AA541941 | AA541941 vJ58d02.Y |
| 145 | 17.2 | 63.7 | 822 | 14 | CE106970 | Shultzomi | 218 | 17 | 63.0 | 617 | 10 | AW701594 | AW701594 uq88B05.Y |
| 146 | 17.2 | 63.7 | 1251 | 28 | BZ579139 | msb2 6216 | 219 | 17 | 63.0 | 618 | 28 | AQ430972 | AQ430972 HS-5048.A |
| 147 | 17.2 | 63.7 | 1430 | 29 | CG755706 | P051-1.H0 | 220 | 17 | 63.0 | 622 | 12 | BG228798 | BG228798 ux61f05.X |
| 148 | 17.2 | 63.0 | 225 | 14 | BH400438 | AG-ND-137 | 221 | 17 | 63.0 | 622 | 13 | BQ418853 | BQ418853 1K58f08.Y |
| 149 | 17.2 | 63.0 | 232 | 28 | W14321 | mb33c04.Y | 222 | 17 | 63.0 | 627 | 29 | CE064146 | CE064146 tigr-gss- |
| 150 | 17.2 | 63.0 | 233 | 29 | CE775929 | tigr-gss- | 223 | 17 | 63.0 | 640 | 13 | BY732910 | BY732910 BY732910 |

224 17 63.0 641 10 BB654211 BB654211
225 17 63.0 647 13 BY737366 BY737366
226 17 63.0 654 14 CB344021 CB344021
227 17 63.0 655 9 AU167947 AU167947
228 17 63.0 657 13 BU653272 BU653272
229 17 63.0 669 29 CE279936 CE279936
230 17 63.0 675 29 CG725184 CG725184
231 17 63.0 676 10 BF482584 BF482584
232 17 63.0 678 9 AU036046 AU036046
233 17 63.0 678 14 CF902666 CF902666
234 17 63.0 679 14 CA268028 CA268028
235 17 63.0 686 13 BU653273 BU653273
236 17 63.0 691 13 CA065477 CA065477
237 17 63.0 698 28 AC281147 AC281147
238 17 63.0 698 29 CC739812 CC739812
239 17 63.0 703 13 BQ204977 BQ204977
240 17 63.0 712 13 BY733199 BY733199
241 17 63.0 714 14 CF731481 CF731481
242 17 63.0 735 12 BI145765 BI145765
243 17 63.0 738 29 AY403448 AY403448
244 17 63.0 759 13 BU847507 BU847507
245 17 63.0 764 12 BI749751 BI749751
246 17 63.0 768 12 BI697648 BI697648
247 17 63.0 776 12 BI330015 BI330015
248 17 63.0 783 10 BF782787 BF782787
249 17 63.0 787 14 CB904653 CB904653
250 17 63.0 787 14 CF875701 CF875701
251 17 63.0 791 12 BG966959 BG966959
252 17 63.0 808 10 BB609684 BB609684
253 17 63.0 828 11 CNS03QYG CNS03QYG
254 17 63.0 828 14 CD575110 CD575110
255 17 63.0 832 12 BI102530 BI102530
256 17 63.0 833 14 CB590660 CB590660
257 17 63.0 833 14 BU558382 BU558382
258 17 63.0 864 14 CB993714 CB993714
259 17 63.0 875 12 BI664826 BI664826
260 17 63.0 878 13 BU557464 BU557464
261 17 63.0 885 11 CNS08KTR CNS08KTR
262 17 63.0 905 14 CB590034 CB590034
263 17 63.0 916 12 BG423488 BG423488
264 17 63.0 949 10 BF781670 BF781670
265 17 63.0 996 12 BM043003 BM043003
266 17 63.0 1010 12 BF162063 BF162063
267 17 63.0 1012 14 CK162997 CK162997
268 17 63.0 1017 13 BY705301 BY705301
269 17 63.0 1070 10 BF134088 BF134088
270 17 63.0 1093 13 BQ889571 BQ889571
271 17 63.0 1110 29 CNS03SFV CNS03SFV
272 17 63.0 1114 10 AW666606 AW666606
273 17 63.0 1127 12 BM478704 BM478704
274 17 63.0 1228 28 BI2058 BI2058
275 17 63.0 1343 9 AA415060 AA415060
276 17 63.0 1501 9 AL930310 AL930310
277 17 63.0 1655 12 BG483774 BG483774
278 17 63.0 1706 13 BU277326 BU277326
279 17 63.0 1720 11 AK005101 AK005101
280 17 63.0 1750 11 AK083308 AK083308
281 17 63.0 1790 11 AK089890 AK089890
282 16.8 62.2 282 10 AW142279 AW142279
283 16.8 62.2 286 10 AW700336 AW700336
284 16.8 62.2 318 12 BQ241421 BQ241421
285 16.8 62.2 389 28 AQ915047 AQ915047
286 16.8 62.2 425 29 CE285999 CE285999
287 16.8 62.2 441 28 BU524840 BU524840
288 16.8 62.2 593 12 AQ696170 AQ696170
289 16.8 62.2 598 12 BU537415 BU537415
290 16.8 62.2 615 12 BJ026424 BJ026424
291 16.8 62.2 648 12 BJ541397 BJ541397
292 16.8 62.2 653 12 BJ531933 BJ531933
293 16.8 62.2 654 12 BJ168369 BJ168369
294 16.8 62.2 702 12 BU584065 BU584065
295 16.8 62.2 730 12 BU507019 BU507019

c 297 16.8 62.2 731 12 BJ165924 BJ165924
298 16.8 62.2 741 12 BJ540699 BJ540699
299 16.8 62.2 780 12 BJ532287 BJ532287
c 300 16.6 61.5 204 9 AA057844 AA057844
ALIGNMENTS
RESULT 1
BB504590
LOCUS BB504590 324 bp mRNA linear EST 26-JUL-2000
DEFINITION BB504590 RIKEN full-length enriched, 0 day neonate kidney Mus
musculus cDNA clone D630049E11 3', mRNA sequence.
ACCESSION BB504590
VERSION BB504590.1 GI:9493384
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 324)
REFERENCE
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Iyawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suganara,Y., Suzuki,H., Tagawa,A.,
Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
Watanabe,S., Yamamura,T., Yamakawa,I., Yano,R., Yasunishi,A.,
Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
CONTACT: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermostabilization of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Itoh,M., Katsunai,T., Katsunai,Y., Ozawa,Y., Muramatsu,M.,
Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
Location/Qualifiers
1..324
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="D630049E11"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/note="Site_1: Sall; Site_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGCGGCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGTCTCGAGTTTCTTTTCTTTT 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 74.8%; Score 20.2; DB 10; Length 324;
Best Local Similarity 88.0%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCT 25
DB 36 CACCGCGCTTTAGGATCGACACCT 60

RESULT 2
LOCUS CA055313/c 630 bp mRNA linear EST 04-MAR-2003
DEFINITION ssalrga510191 mixed_tissue Salmo salar cDNA, mRNA sequence.
ACCESSION CA055313
VERSION CA055313.1 GI:24385556
KEYWORDS EST.
SOURCE 'Salmo salar (Atlantic salmon)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE 1 (bases 1 to 630)
AUTHORS GRASP Consortium, Davidson, W.S., Koop, B.F. and
TITLE A survey of Salmo salar transcripts from high complexity cDNA
JOURNAL Unpublished (2002)
COMMENT Contact: Koop BF
Centre for Biomedical Research
University of Victoria
PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
Tel: 250 472 4067
Fax: 250 472 4075
Email: bkoop@uvic.ca
Genome Sciences Centre, BC Cancer Agency
cDNA preparation, sequencing and bioinformatics:
S Lee, T Olson, P Pandoh, A Prabhu, D Smalhus, L Spence, J Stott,
S Taylor, G Yang, J Schein, S Jones and M Marra.

Location/Qualifiers

1. .630
/organism="Salmo salar"
/mol_type="mRNA"
/strain="McConnell"
/db_xref="taxon:8030"
/clone_lib="mixed tissue"
/note="Vector: pCMVSPORT6; Library Creator: Research Genetics; Atlantic salmon tissue contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton, B.C.), Rachel Roper (University of Victoria)"

ORIGIN

Query Match 72.6%; Score 19.6; DB 13; Length 630;
Best Local Similarity 84.8%; Pred. No. 4.6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
DB 348 ATCTAGCTTTAGCATCGACACCTGA 323

RESULT 3

LOCUS AQ259762/c 634 bp DNA linear GSS 23-OCT-1998

DEFINITION nxb0023D22f CUGI Rice BAC Library Oryza sativa (japonica cultivar-group) genomic clone nxb0023D22f, genomic survey sequence.

ACCESSION AQ259762
VERSION AQ259762.1 GI:3784244
KEYWORDS GSS.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 634)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence stop: 273.

Location/Qualifiers

1. .634
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nxb0023D22f"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice BAC Library"
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

ORIGIN

Query Match 71.9%; Score 19.4; DB 28; Length 634;
Best Local Similarity 95.2%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

QY 6 CCGCTTTAGGATCGACACCTG 26
DB 294 GCCTTTAGGATCGACACCTG 274


```

RESULT 4
BQ683442
LOCUS
DEFINITION
  BQ683442 1130 bp mRNA linear EST 15-JUL-2002
  5', mRNA sequence.
ACCESSION
  BQ683442
VERSION
  BQ683442.1 GI:21795121
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 1130)
  NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS
  National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: DCTD/DPF
  CDNA Library Preparation: Rubin Laboratory
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLCW2432 row: k column: 23
  High quality sequence start: 14
  High quality sequence stop: 246.
FEATURES
  source
    location/Qualifiers
    1..1130
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:6265486"
    /tissue_type="melanotic melanoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH MGC 112"
    /note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:
    EcoRI; cDNA made by oligo-dT priming. Directionally cloned
    into EcoRI/XhoI sites using the following 5' adaptor:
    GGCACGAG(G). Library constructed by Ling Hong in the
    laboratory of Gerald M. Rubin (University of California,
    Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
    Superscript II RT (Life Technologies). Note: this is a
    NIH_MGC Library."
ORIGIN
  Query Match 71.1%; Score 19,2; DB 13; Length 1130;
  Best Local Similarity 87.5%; Pred. No. 6e+02; Mismatches 0; Gaps 0;
  Matches 21; Conservative 0; Indels 3; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACAC 24
|||||
Db 1106 CACCGCGCTTTAGGACCCACCC 1129

RESULT 5
AZ143037/c
LOCUS
DEFINITION
  AZ143037 483 bp DNA linear GSS 28-AUG-2000
  SP 0041 B1 D04 SP6E Strongylocentrotus purpuratus, purple sea
  urchin, sperm genomic BAC library Strongylocentrotus purpuratus
  genomic clone Plate-41 Col=7 Row=H, genomic survey sequence.
ACCESSION
  AZ143037
VERSION
  AZ143037.1 GI:8294940
KEYWORDS
  GSS.
SOURCE
  Strongylocentrotus purpuratus
  ORGANISM
    Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
    Echinoidea; Euechinoidea; Echinacea; Echinoida;
    Strongylocentrotidae; Strongylocentrotus.
  1 (bases 1 to 483)
  Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
  Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
  Wray,G.A., Ettensom,C.A., Lehrach,H., Britten,R.J., Davidson,E.H.
  A sea urchin genome project: Sequence scan, virtual map, and
  additional resources
  Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
10920195
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plats: 41 row: H column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 483.
FEATURES
  source
    location/Qualifiers
    1..483
    /organism="Strongylocentrotus purpuratus"
    /mol_type="genomic DNA"
    /db_xref="taxon:7668"
    /clone_lib="Plate=41 Col=7 Row=H"
    /clone_lib="Strongylocentrotus purpuratus, purple sea
    urchin, sperm genomic BAC library"
    /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
    DH10B"
ORIGIN
  Query Match 70.4%; Score 19; DB 28; Length 483;
  Best Local Similarity 81.5%; Pred. No. 7.8e+02;
  Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTCA 27
|||||
Db 271 CATCGCGCTTTAAGCAGCAACCTCA 245

RESULT 6
CE809540
LOCUS
DEFINITION
  tigr-gss-dog-1700033511007 Dog Library Canis familiaris genomic,
  genomic survey sequence.
ACCESSION
  CE809540
VERSION
  CE809540.1 GI:37150524
KEYWORDS
  GSS.
SOURCE
  Canis familiaris (dog)
  ORGANISM
    Canis familiaris
    Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
  1 (bases 1 to 533)
  Kirkness,E.F., Batna,V., Halpern,A.L., Levy,S., Remington,K.,
  Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
  Venter,J.C.
  The dog genome: survey sequencing and comparative analysis
  Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES
  source
    location/Qualifiers
    1..533
    /organism="Canis familiaris"
    /mol_type="genomic DNA"
    /strain="Standard Poodle"

```

```

Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,
Wray,G.A., Ettensom,C.A., Lehrach,H., Britten,R.J., Davidson,E.H.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plats: 41 row: H column: 7
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 483.
FEATURES
  source
    location/Qualifiers
    1..483
    /organism="Strongylocentrotus purpuratus"
    /mol_type="genomic DNA"
    /db_xref="taxon:7668"
    /clone_lib="Plate=41 Col=7 Row=H"
    /clone_lib="Strongylocentrotus purpuratus, purple sea
    urchin, sperm genomic BAC library"
    /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
    DH10B"
ORIGIN
  Query Match 70.4%; Score 19; DB 28; Length 483;
  Best Local Similarity 81.5%; Pred. No. 7.8e+02;
  Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTTAGGATCGACACCTCA 27
|||||
Db 271 CATCGCGCTTTAAGCAGCAACCTCA 245

RESULT 6
CE809540
LOCUS
DEFINITION
  tigr-gss-dog-1700033511007 Dog Library Canis familiaris genomic,
  genomic survey sequence.
ACCESSION
  CE809540
VERSION
  CE809540.1 GI:37150524
KEYWORDS
  GSS.
SOURCE
  Canis familiaris (dog)
  ORGANISM
    Canis familiaris
    Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
  1 (bases 1 to 533)
  Kirkness,E.F., Batna,V., Halpern,A.L., Levy,S., Remington,K.,
  Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
  Venter,J.C.
  The dog genome: survey sequencing and comparative analysis
  Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES
  source
    location/Qualifiers
    1..533
    /organism="Canis familiaris"
    /mol_type="genomic DNA"
    /strain="Standard Poodle"

```



```
RESULT 9
BJ010761
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1.603
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="HG-rR"
/db_xref="taxon:8090"
/clone="MF01SSA149B11"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stages="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"
ORIGIN
Query Match 68.9%; Score 18.6; DB 12; Length 603;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 3 CCCGCGCTTTAGGATCGACACTGA 27
Db 261 CCCGACATTGAGATCGACACTGA 285
RESULT 10
BG500183/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
662 bp mRNA linear EST 27-MAR-2001
602547620F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4669877 5',
mRNA sequence.
BG500183
BG500183.1 GI:13461700
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: DCTD/DRP
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCm482 row: h column: 06
High quality sequence stop: 299.
Location/Qualifiers
1..662
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4669877"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC 60"
/notes="Organ: prostate; Vector: pDNR-LTB (Clontech);
Site 1: SfiI (ggccctcgcc); Site 2: SfiI
(ggcattatggc); Double-stranded cDNA was prepared from
cell line RNA; 5' and 3' adaptors were used in cloning as
follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and
3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.5
kb (range 0.9-4.0 kb). 14/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
ORIGIN
Query Match 68.9%; Score 18.6; DB 12; Length 662;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 ACCCGCGCTTTAGGATCGACACTG 26
Db 429 ACACAGCTTTAGGAGCCACACTG 405
RESULT 11
CNS08HL2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
719 bp mRNA linear HTC 07-JAN-2003
Single read from an extremity of a full-length cDNA clone made from
Anopheles gambiae total adult females. 3-PRIME end of clone
FK0AAA19BG04 of strain 6-9 of Anopheles gambiae (African malaria
mosquito).
BX012210
BX012210.1 GI:27561430
HTC.
Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 719)
Genoscope.
Direct Submission
Submitted (06-JAN-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Location/Qualifiers
1..719
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="6-9"
/db_xref="taxon:7165"
/clone="FK0AAA19BG04"
/plasmid="pME18S-FL"
/notes="end : 3-PRIME"
ORIGIN
Query Match 68.9%; Score 18.6; DB 11; Length 719;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 3 CCCGCGCTTTAGGATCGACACTGA 27
```

Db 325 CCGCGCTTTACGGTCGACGCTTA 349

CE802854 tigr-gss-dog-17000331449669 Dog Library Canis familiaris genomic,
genomic survey sequence. GSS 30-SEP-2003

CE802854 757 bp DNA linear
CE802854.1 GI:37143776

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 757)
Kirkness,E.F., Batina,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,M., Fraser,C.M. and
Venter,J.C.

REFERENCE
AUTHORS
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)

TITLE
JOURNAL
MEDLINE
PUBMED
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..757
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 68.9%; Score 18.6; DB 29; Length 757;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGCGCTTTAGGATCGACACTGA 27
|||||
Db 474 CCGGACCTTAGAATCGACACTGA 450
|||||

RESULT 13
BF789780/c
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 922)
NIH-MGC <http://mgs.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA library preparation: Life Technologies, Inc.
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment

FEATURES
source
Location/Qualifiers
1..1131
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
1..1131
/locus_tag="HCM1653"

ORIGIN
Query Match 68.9%; Score 18.6; DB 29; Length 1131;
Best Local Similarity 84.0%; Pred. No. 1.5e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9811 row: n column: 07
High quality sequence stop: 652.

FEATURES
source
Location/Qualifiers
1..922
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI CGAP Kid14"
/lab_host="DR10B (Ti phage-resistant)"
/notes="Organ: Kidney; Vector: pQWV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

ORIGIN
Query Match 68.9%; Score 18.6; DB 10; Length 922;
Best Local Similarity 84.0%; Pred. No. 1.4e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCGGCTTTAGGATCGACACTG 26
|||||
Db 895 ACCCGCTTTAGGAACGCCCTG 871
|||||

RESULT 14
AY403727
LOCUS
DEFINITION
Mus musculus HCM1653 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1131)
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Farrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Farrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment

FEATURES
source
Location/Qualifiers
1..1131
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
1..1131
/locus_tag="HCM1653"

ORIGIN
Query Match 68.9%; Score 18.6; DB 29; Length 1131;
Best Local Similarity 84.0%; Pred. No. 1.5e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

| | |
|-------------------|--|
| <p>QY</p> | 3 CCGGCGCTTTAGATGCACACTGA 27 922 CCGGGCCTTCAGGTAGACTTCTGTA 946 |
| <p>Db</p> | |
| <p>RESULT 15</p> | |
| <p>LOCUS</p> | AK006855 1342 bp mRNA linear HTC 20-SEP-2003 |
| <p>DEFINITION</p> | Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:170061J02 product:hypothetical Actin and actin-like containing protein, full insert sequence. |
| <p>ACCESSION</p> | AK006855 |
| <p>VERSION</p> | AK006855.1 GI:12840133 |
| <p>KEYWORDS</p> | HTC; CAP trapper; |
| <p>SOURCE</p> | Mus musculus (house mouse) |
| <p>ORGANISM</p> | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. |
| <p>REFERENCE</p> | |
| <p>AUTHORS</p> | Carninci, P. and Hayashizaki, Y. |
| <p>TITLE</p> | High-efficiency full-length cDNA cloning |
| <p>JOURNAL</p> | Meth. Enzymol. 303, 19-44 (1999) |
| <p>MEDLINE</p> | 99279253 |
| <p>PUBMED</p> | 10349636 |
| <p>REFERENCE</p> | |
| <p>AUTHORS</p> | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. |
| <p>TITLE</p> | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes |
| <p>JOURNAL</p> | Genome Res. 10 (10), 1617-1630 (2000) |
| <p>MEDLINE</p> | 20499374 |
| <p>PUBMED</p> | 11042159 |
| <p>REFERENCE</p> | |
| <p>AUTHORS</p> | Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoch, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoshida, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. |
| <p>TITLE</p> | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipicillary sequencer |
| <p>JOURNAL</p> | Genome Res. 10 (11), 1757-1771 (2000) |
| <p>MEDLINE</p> | 20530913 |
| <p>PUBMED</p> | 11076861 |
| <p>REFERENCE</p> | |
| <p>AUTHORS</p> | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. |
| <p>TITLE</p> | Functional annotation of a full-length mouse cDNA collection |
| <p>JOURNAL</p> | Nature 409, 685-690 (2001) |
| <p>REFERENCE</p> | |
| <p>AUTHORS</p> | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. |
| <p>TITLE</p> | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs |
| <p>JOURNAL</p> | Nature 420, 563-573 (2002) |
| <p>REFERENCE</p> | |
| <p>AUTHORS</p> | Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoch, Y., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komoi, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nunazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Saigo, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Teijma, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. |
| <p>TITLE</p> | Direct Submission |
| <p>JOURNAL</p> | Submitted (10-JULI-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC). |

Eukaryota; Metazoa; Chromadorea; Rhabditida; Strongylidae; Ancylostomatoidea; Ancylostomatidae; Bunostominae; Necator. 1 (bases 1 to 257)

REFERENCE
AUTHORS
Blaxter M.L., Parkinson J., Whitten, C., Daub, J., Guiliano, D., Hall N., Quayle, M. and Barrrell, B.

TITLE
JOURNAL
COMMENT
Edinburgh University/Sanger Centre Nematode EST Project Unpublished (2000)

Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.

Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

The library was prepared for Dr. David Pritchard University of Nottingham. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail and Bart Barrrell). The sequence contained a PolyA tail (trimmed)

PCR Primers
FORWARD: SAC
BACKWARD: T7PL
Plate: 39 row: H column: 07
Seq primer: SAC
High quality sequence stop: 257.

Location/Qualifiers
1. .257
/organism="Necator americanus"
/mol_type="mRNA"
/db_xref="taxon:51031"
/clone="Na L3_39H07"
/sex="Mixed"
/dev_stage="L3"
/clone_lib="Necator americanus (parasitic nematode) L3"
/notes="Vector: pCMV-PCR vector; Site 1: EcoRI (5'end); Site 2: XhoI (3'end); Necator americanus is a human hookworm, responsible for debilitating anaemia. The library was constructed by David Pritchard (University of Nottingham, UK) from mRNA from N.americanus L3 stage (Nottingham strain) maintained in hamsters."

ORIGIN

Query Match 67.4%; Score 18.2; DB 13; Length 257;
Best Local Similarity 87.0%; Pred. No. 1.5e-03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGCTTTAGGATCGACACTGA 27
|||||
DDB 179 CACGCTCTAGGATCGCACCTGA 157
|||||

RESULT 17
EBE029790/c
LOCUS
Kp34h02.y1 TEN95TW-SFHH Strongyloides stercoralis cDNA 5' similar to TR:061855 O61855 F33E11.4 PROTEIN. i, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Strongyloides stercoralis
Strongyloides stercoralis
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Paragrolaimoidea; Strongyloidea; Strongyloides.

REFERENCE
AUTHORS
McCarte J., Clifton S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Darte, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarisvilli, R., Ronko, I., Kennedy, S., Meguire, D., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE
JOURNAL
COMMENT
The Washington Univ. Nematode EST project, 1999
Unpublished (1999)

web page at: <http://www.resgen.com/> Please reference the id listed below when ordering this clone: Source lab clone id -
Seq primer: -40RP from Gibco
High quality sequence stop: 458.

FEATURES
source

Location/Qualifiers
1..542
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XENOPUS SOURCE ID:"
/tissue_type="oocyte (stages 5 and 6)"
/lab_host="Top-10 F"
/clone_lib="Xenopus laevis oocyte"
/note="Vector: pBluescript SK⁺; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2ug of poly A⁺ RNA. EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F⁺. Original library construction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Normalized by Jihwan Song (Song, Cho and Blumberg, unpublished). Note: This is a Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 67.4%; Score 18.2; DB 10; Length 542;
Best Local Similarity 87.0%; Pred. No. 1.8e+03; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 3;

Qy 5 CGCGCTTTAGGATCGACACCTGA 27
|||||
Db 390 CGCGCTTGAGGATCTCCACCTGA 368

RESULT 19
CA164304

LOCUS
DEFINITION
SCSBR23119C04.g RZ3 Saccharum officinarum cDNA clone SCSBR23119C04
5', mRNA sequence.

ACCESSION
CA164304
CA164304.1 GI:35081583

KEYWORDS
EST.
SOURCE
Saccharum officinarum
Saccharum officinarum

ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.

REFERENCE
1 (bases 1 to 802)
Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.

AUTHORS
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

JOURNAL
Contact: Arruda P

COMMENT
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137

Fax: 55 19 3788 1089
Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at

<http://www.bcccenter.fcav.unesp.br>

Plate: 119 row: C column: 04
Seq primer: T7 Promoter Primer.

Location/Qualifiers
1..802

/organism="Saccharum officinarum"
/mol_type="mRNA"

/db_xref="taxon:4547"
/clone="SCSBR23119C04"
/lab_host="DH10B"
/clone_lib="RZ3"

/note="Organ: Shoot-root transition zone from adult plants; Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [shoot-root transition zone from adult plants]. cDNA was prepared from polyA⁺ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucet.lad.ic.unicamp.br/public>"

ORIGIN

Query Match 67.4%; Score 18.2; DB 13; Length 802;
Best Local Similarity 87.0%; Pred. No. 2e+03; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 3;

Qy 4 CGCGCTTTAGGATCGACACCTG 26
|||||
Db 777 CGCGCTTTAGGCGCGACACCTG 799

RESULT 20
CNS055CF

LOCUS
DEFINITION

CNS055CF 1101 bp DNA linear GSS 26-JUL-2000
Tetraodon nigroviridis genome survey sequence T3 end of clone
030A22 of library A from Tetraodon nigroviridis, genomic survey

sequence.
ACCESSION
AL321864.1 GI:9554748

VERSION
GSS; genome survey sequence.
KEYWORDS
Tetraodon nigroviridis

SOURCE
Tetraodon nigroviridis

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.

Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)

20296633
PUBMED
10835645

REFERENCE
AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

Genome Res. 10 (7), 939-949 (2000)

20359837
PUBMED
10839143

3 (bases 1 to 1101)
Genoscope.

Direct Submission
TITLE
Submitted (12-APR-2000) Genoscope - Centre National de Sequençage :

Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at

<http://www.genoscope.cns.fr/Tetraodon>.

Location/Qualifiers
1..1101

/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"


```

/db_xref="taxon:99883"
/clone="030A22"
/clone_lib="A"
/note="Genoscope sequence ID : COAA030B11A1-end : T3"

ORIGIN

Query Match      67.4%; Score 18.2; DB 29; Length 1101;
Best Local Similarity 83.3%; Pred. NO. 2.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CCGCGCTTTAGGATCGACACCTG 26
    |||||
Db 848 CCGCTCTTTGGATCAGACCTG 871

RESULT 21
BB563010/c
LOCUS
DEFINITION
BB563010 RIKEN full-length enriched, adult male heart Mus musculus
CDNA clone 1010001L20 5', mRNA sequence.
BB563010
BB563010.1 GI:11453902
EST.
Mus musculus (house mouse)
Mus musculus
Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 286)
Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2): 520-524 (1998)
Itoh,M., Kiteunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Okazaki,Y., Muramatsu,M.,
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5): 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. .286
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="1010001L20"
/sex="male"
/tissue_type="heart"
/dev_stage="adult"

FEATURES
source

```

```

/lab_host="BNV132"
/clone_lib="RIKEN full-length enriched, adult male heart"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGGCACTCGAGTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGAGATCCAGAGCTCATTAATTAAACCCCCCCCCC 3'].
cDNA was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match      66.7%; Score 18; DB 10; Length 286;
Best Local Similarity 80.8%; Pred. No. 1.9e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTGA 27
    |||||
Db 40 AGCGCGCTTTAGGACCTTCACGAGA 15

RESULT 22
AQ557945
LOCUS
DEFINITION
HS 2067 B1.F06.T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2067 Col=11 Row=L, genomic survey
sequence.
AQ557945
AQ557945.1 GI:4917677
GSS.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 457)
Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
93380589
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com)
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2067 row: L column: 11
Seq primer: T7
Class: BAC ends
High quality sequence stop: 457.
Location/Qualifiers
1. .457
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2067 Col=11 Row=L"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

FEATURES
source

```


ORIGIN

Query Match 66.7%; Score 18; DB 28; Length 457;
 Best Local Similarity 80.8%; Pred. No. 2.1e+03;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTAGGATCGACACCTG 26

DB 274 CACCGCGCTTACAGATCCACCTG 299

RESULT 23

AZ665932/c

LOCUS

DEFINITION 1M0547H12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0547H12 R, genomic survey sequence.

ACCESSION AZ665932

VERSION AZ665932.1

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (Bases 1 to 490)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112 USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: rdunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0547 row: H column: 12

Seq primer: CACACGAGAAACAGCTATGACC

Class: Plasmid ends

High quality sequence stop: 490.

FEATURES

source

1. .490
 /location=Qualifiers
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clones="UUGC1M0547H12"
 /sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 66.7%; Score 18; DB 28; Length 490;
 Best Local Similarity 80.8%; Pred. No. 2.2e+03;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCGCGCTTAGGATCGACACCTG 26

DB 208 CACCGCGCTTAGGCTCCACACATG 183

RESULT 24

EG851071

LOCUS

DEFINITION 1024030B12.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION EG851071

VERSION EG851071.1

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

REFERENCE 1 (bases 1 to 506)
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

AUTHORS

Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P., McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; project phase 2

JOURNAL

COMMENT

Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

1. .506
 /organism="Chlamydomonas reinhardtii"

/mol_type="mRNA"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2.

polyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.

pluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN

Query Match 66.7%; Score 18; DB 12; Length 506;
 Best Local Similarity 80.8%; Pred. No. 2.2e+03;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTAGGATCGACACCTGA 27

DB 205 ACACGAGCTTTGGACGTCACCTGA 230

RESULT 25

CA260332

LOCUS

DEFINITION SCLRT3034E08.g RT3 Saccharum officinarum cDNA clone SCLRT3034E08

CA260332 612 bp mRNA linear EST 26-SEP-2003

/clone="tcbk0045c.a.11"
 /tissue_type="multi-tissues"
 /dev_stage="from embryos to adults"
 /lab_host="DH10B"
 /clone_lib="AGENAE Rainbow trout multi-tissues-normalized (tcbk)"
 /note="vector: pT73D-pac; Rainbow trout multi-tissues-normalized + 2 subtractions; Clone distribution : AGENAE Resource centre. Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73"

ORIGIN

Query Match 66.7%; Score 18; DB 13; Length 763;

Best Local Similarity 80.8%; Pred. No. 2.4e+03;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACCTGA 27

Db 435 ATCTAGCTTTAGCATCGCCACCTGA 460

RESULT 28

BZ704611

LOCUS BZ704611 771 bp DNA linear GSS 19-FEB-2003
 DEFINITION PUCF02TD ZM_0.6_1.0_KB Zea mays genomic clone ZMBBta126A04,
 genomic survey sequence.

ACCESSION BZ704611

VERSION BZ704611.1 GI:28425019

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 771)

Whitelaw C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and

Bennetzen, J.

Maize Genomics Consortium

Unpublished (2003)

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@cigr.org

Seq primer: Tg

Class: sheared ends.

Location/Qualifiers

1..771

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMBBta126A04"

/clone_lib="ZM 0.6_1.0_KB"

/note="vector: PCR4-TOPO; site 1: EcoRI; 0.6-1.0 kb high

COT selected genomic DNA library"

ORIGIN

Query Match 66.7%; Score 18; DB 28; Length 771;

Best Local Similarity 80.8%; Pred. No. 2.4e+03;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CACCGCGCTTTAGGATCGACACCTG 26

Db 230 CACCGCGGATTAGATAGAAACCTG 255

RESULT 29

BX872970

LOCUS

DEFINITION BX872970 AGENAE Rainbow trout multi-tissues-normalized (tcbk)
 Oncorhynchus mykiss cDNA clone tcbk0025c.d.13 5prim, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Oncorhynchus mykiss (rainbow trout)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 1 (bases 1 to 777)

REFERENCE

Govoroun, M., Guiguen, Y. and Le Gac, F.

Construction and primary characterization of normalized cDNA

libraries in rainbow trout, *Oncorhynchus mykiss*

Unpublished (2003)

JOURNAL

COMMENT

Contact: Guiguen Y

INRA - SCRIBE

Campus de Beaulieu, RENNES cedex, 35042, France

Tel: 02.23.48.50.09

Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us

at sigenasupport@jouy.inra.fr to obtain the Chromatogram of this

sequence.

Plate: 0025 row: d column: 13

Seq primer: M13R.

FEATURES

source

1..777

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:8022"

/clone="tcbk0025c.d.13"

/tissue_type="multi-tissues"

/dev_stage="from embryos to adults"

/lab_host="DH10B"

/clone_lib="AGENAE Rainbow trout multi-tissues-normalized

(tcbk)"

/note="vector: pT73D-pac; Rainbow trout

multi-tissues-normalized + 2 subtractions; Clone

distribution : AGENAE Resource centre. Francois PIUMI,

Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et

Etude du genome (LREG), Domaine de Vilvert, 78352,

Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33

(0) 1.34.65.22.73"

Query Match 66.7%; Score 18; DB 13; Length 777;

Best Local Similarity 80.8%; Pred. No. 2.4e+03;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACCCGCGCTTTAGGATCGACACCTGA 27

Db 582 ATCTAGCTTTAGCATCGCCACCTGA 607

RESULT 30

BX876110

LOCUS

DEFINITION BX876110 AGENAE Rainbow trout multi-tissues-normalized (tcbk)
 Oncorhynchus mykiss cDNA clone tcbk0014c.f.23 5prim, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Oncorhynchus mykiss (rainbow trout)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 1 (bases 1 to 777)

REFERENCE

Govoroun, M., Guiguen, Y. and Le Gac, F.

Construction and primary characterization of normalized cDNA

libraries in rainbow trout, Oncorhynchus mykiss
Unpublished (2003)
Contact: Guiguen Y
INRA - SCRIBE
Campus de beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at signasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0014 row: f column: 23
Seq primer: M13R.

FEATURES
source
1..777
Location/Qualifiers
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcBK0014C.f.23"
/tissue_type="multi-tissues"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="AGENAE Rainbow trout multi-tissues-normalized
(tcBK)"
/note="Vector: pT73D-pac; Rainbow trout
multi-tissues-normalized + 2 subtractions; Clone
distribution : AGENAE Resource centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

ORIGIN

Query Match 66.7%; Score 18; DB 13; Length 777;
Best Local Similarity 80.8%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTAGGATCGACACCTGA 27
Db 582 ATCTAGCTTTAGCATGCCACCTGA 607

Search completed: June 20, 2004, 14:13:35
Job time : 1288.71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 152.714 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-12
Perfect score: 27
Sequence: 1 caccgcgcttagatcgacacctga 27

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3017436 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034952

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| 1 | 19.2 | 71.1 | 2940917 | 13 | US-10-027-632-174763 |
| 2 | 19.2 | 71.1 | 2940917 | 16 | US-10-027-632-174763 |
| 3 | 18.6 | 68.9 | 2000 | 9 | US-09-938-842A-4597 |
| 4 | 18.6 | 68.9 | 2000 | 11 | US-09-938-842A-4597 |
| 5 | 18.6 | 68.9 | 2000 | 11 | US-09-938-842A-4597 |
| 6 | 18.6 | 68.9 | 2000 | 11 | US-09-938-842A-4597 |
| 7 | 18.6 | 68.9 | 2000 | 11 | US-09-938-842A-4597 |
| 8 | 18.6 | 68.9 | 2000 | 11 | US-09-938-842A-4597 |
| 9 | 17.6 | 65.2 | 894 | 15 | US-10-138-970A-1 |
| 10 | 17.6 | 65.2 | 894 | 15 | US-10-138-970A-1 |
| 11 | 17.6 | 65.2 | 894 | 15 | US-10-138-970A-1 |
| 12 | 17.6 | 65.2 | 894 | 15 | US-10-138-970A-1 |
| 13 | 17.4 | 64.4 | 842 | 13 | US-10-027-632-324925 |
| 14 | 17.4 | 64.4 | 842 | 13 | US-10-027-632-324925 |

| | | | | | | |
|------|------|------|----------------------|-----|----|--------------------|
| 64.4 | 17.4 | C 15 | US-10-027-632-324925 | 842 | 16 | Sequence 324925, |
| 64.4 | 17.4 | C 16 | US-10-027-632-325019 | 842 | 16 | Sequence 325019, |
| 64.4 | 17.4 | C 17 | US-10-027-632-325019 | 842 | 16 | Sequence 27104, A |
| 64.4 | 17.4 | C 18 | US-10-027-632-325019 | 842 | 16 | Sequence 103399, |
| 64.4 | 17.4 | C 19 | US-10-027-632-325019 | 842 | 16 | Sequence 6630, Ap |
| 64.4 | 17.4 | C 20 | US-10-027-632-325019 | 842 | 16 | Sequence 26417, A |
| 64.4 | 17.4 | C 21 | US-10-027-632-325019 | 842 | 16 | Sequence 135, App |
| 64.4 | 17.4 | C 22 | US-10-027-632-325019 | 842 | 16 | Sequence 28726, A |
| 64.4 | 17.4 | C 23 | US-10-027-632-325019 | 842 | 16 | Sequence 1937, Ap |
| 64.4 | 17.4 | C 24 | US-10-027-632-325019 | 842 | 16 | Sequence 2319, A |
| 64.4 | 17.4 | C 25 | US-10-027-632-325019 | 842 | 16 | Sequence 45829, A |
| 64.4 | 17.4 | C 26 | US-10-027-632-325019 | 842 | 16 | Sequence 30891, A |
| 64.4 | 17.4 | C 27 | US-10-027-632-325019 | 842 | 16 | Sequence 131520, |
| 64.4 | 17.4 | C 28 | US-10-027-632-325019 | 842 | 16 | Sequence 661, App |
| 64.4 | 17.4 | C 29 | US-10-027-632-325019 | 842 | 16 | Sequence 108794, |
| 64.4 | 17.4 | C 30 | US-10-027-632-325019 | 842 | 16 | Sequence 14328, A |
| 64.4 | 17.4 | C 31 | US-10-027-632-325019 | 842 | 16 | Sequence 40188, A |
| 64.4 | 17.4 | C 32 | US-10-027-632-325019 | 842 | 16 | Sequence 1494, A |
| 64.4 | 17.4 | C 33 | US-10-027-632-325019 | 842 | 16 | Sequence 8860, Ap |
| 64.4 | 17.4 | C 34 | US-10-027-632-325019 | 842 | 16 | Sequence 24148, A |
| 64.4 | 17.4 | C 35 | US-10-027-632-325019 | 842 | 16 | Sequence 20169, A |
| 64.4 | 17.4 | C 36 | US-10-027-632-325019 | 842 | 16 | Sequence 43341, A |
| 64.4 | 17.4 | C 37 | US-10-027-632-325019 | 842 | 16 | Sequence 156, App |
| 64.4 | 17.4 | C 38 | US-10-027-632-325019 | 842 | 16 | Sequence 8437, Ap |
| 64.4 | 17.4 | C 39 | US-10-027-632-325019 | 842 | 16 | Sequence 8101, Ap |
| 64.4 | 17.4 | C 40 | US-10-027-632-325019 | 842 | 16 | Sequence 8101, Ap |
| 64.4 | 17.4 | C 41 | US-10-027-632-325019 | 842 | 16 | Sequence 9001, Ap |
| 64.4 | 17.4 | C 42 | US-10-027-632-325019 | 842 | 16 | Sequence 9001, Ap |
| 64.4 | 17.4 | C 43 | US-10-027-632-325019 | 842 | 16 | Sequence 259, App |
| 64.4 | 17.4 | C 44 | US-10-027-632-325019 | 842 | 16 | Sequence 460, App |
| 64.4 | 17.4 | C 45 | US-10-027-632-325019 | 842 | 16 | Sequence 687, App |
| 64.4 | 17.4 | C 46 | US-10-027-632-325019 | 842 | 16 | Sequence 687, App |
| 64.4 | 17.4 | C 47 | US-10-027-632-325019 | 842 | 16 | Sequence 28094, A |
| 64.4 | 17.4 | C 48 | US-10-027-632-325019 | 842 | 16 | Sequence 26133, A |
| 64.4 | 17.4 | C 49 | US-10-027-632-325019 | 842 | 16 | Sequence 35426, A |
| 64.4 | 17.4 | C 50 | US-10-027-632-325019 | 842 | 16 | Sequence 41735, A |
| 64.4 | 17.4 | C 51 | US-10-027-632-325019 | 842 | 16 | Sequence 63, Appl |
| 64.4 | 17.4 | C 52 | US-10-027-632-325019 | 842 | 16 | Sequence 37320, A |
| 64.4 | 17.4 | C 53 | US-10-027-632-325019 | 842 | 16 | Sequence 39316, A |
| 64.4 | 17.4 | C 54 | US-10-027-632-325019 | 842 | 16 | Sequence 11722, A |
| 64.4 | 17.4 | C 55 | US-10-027-632-325019 | 842 | 16 | Sequence 115, App |
| 64.4 | 17.4 | C 56 | US-10-027-632-325019 | 842 | 16 | Sequence 78621, A |
| 64.4 | 17.4 | C 57 | US-10-027-632-325019 | 842 | 16 | Sequence 268282, |
| 64.4 | 17.4 | C 58 | US-10-027-632-325019 | 842 | 16 | Sequence 268282, |
| 64.4 | 17.4 | C 59 | US-10-027-632-325019 | 842 | 16 | Sequence 19476, A |
| 64.4 | 17.4 | C 60 | US-10-027-632-325019 | 842 | 16 | Sequence 1537, Ap |
| 64.4 | 17.4 | C 61 | US-10-027-632-325019 | 842 | 16 | Sequence 1516, Ap |
| 64.4 | 17.4 | C 62 | US-10-027-632-325019 | 842 | 16 | Sequence 5883, Ap |
| 64.4 | 17.4 | C 63 | US-10-027-632-325019 | 842 | 16 | Sequence 9323, Ap |
| 64.4 | 17.4 | C 64 | US-10-027-632-325019 | 842 | 16 | Sequence 1574, Ap |
| 64.4 | 17.4 | C 65 | US-10-027-632-325019 | 842 | 16 | Sequence 4757, A |
| 64.4 | 17.4 | C 66 | US-10-027-632-325019 | 842 | 16 | Sequence 39509, A |
| 64.4 | 17.4 | C 67 | US-10-027-632-325019 | 842 | 16 | Sequence 36877, A |
| 64.4 | 17.4 | C 68 | US-10-027-632-325019 | 842 | 16 | Sequence 19724, A |
| 64.4 | 17.4 | C 69 | US-10-027-632-325019 | 842 | 16 | Sequence 26318, A |
| 64.4 | 17.4 | C 70 | US-10-027-632-325019 | 842 | 16 | Sequence 15461, A |
| 64.4 | 17.4 | C 71 | US-10-027-632-325019 | 842 | 16 | Sequence 32483, A |
| 64.4 | 17.4 | C 72 | US-10-027-632-325019 | 842 | 16 | Sequence 1601, Ap |
| 64.4 | 17.4 | C 73 | US-10-027-632-325019 | 842 | 16 | Sequence 1520, Ap |
| 64.4 | 17.4 | C 74 | US-10-027-632-325019 | 842 | 16 | Sequence 6112, Ap |
| 64.4 | 17.4 | C 75 | US-10-027-632-325019 | 842 | 16 | Sequence 12781, A |
| 64.4 | 17.4 | C 76 | US-10-027-632-325019 | 842 | 16 | Sequence 1342, Ap |
| 64.4 | 17.4 | C 77 | US-10-027-632-325019 | 842 | 16 | Sequence 38, Appl |
| 64.4 | 17.4 | C 78 | US-10-027-632-325019 | 842 | 16 | Sequence 3707, Ap |
| 64.4 | 17.4 | C 79 | US-10-027-632-325019 | 842 | 16 | Sequence 72, Appl |
| 64.4 | 17.4 | C 80 | US-10-027-632-325019 | 842 | 16 | Sequence 393, App |
| 64.4 | 17.4 | C 81 | US-10-027-632-325019 | 842 | 16 | Sequence 3, Appl |
| 64.4 | 17.4 | C 82 | US-10-027-632-325019 | 842 | 16 | Sequence 5, Appl |
| 64.4 | 17.4 | C 83 | US-10-027-632-325019 | 842 | 16 | Sequence 1, Appl |
| 64.4 | 17.4 | C 84 | US-10-027-632-325019 | 842 | 16 | Sequence 1827, App |
| 64.4 | 17.4 | C 85 | US-10-027-632-325019 | 842 | 16 | Sequence 1867, Ap |
| 64.4 | 17.4 | C 86 | US-10-027-632-325019 | 842 | 16 | Sequence 97, Appl |
| 64.4 | 17.4 | C 87 | US-10-027-632-325019 | 842 | 16 | Sequence 325019, |


```
; LENGTH: 2940917
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2940917)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174763

Query Match
Best Local Similarity 71.1%; Score 19.2; DB 16; Length 2940917;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACC 24
   ||||| ||||| ||||| |||||
Db 1793893 CACCCGCGCTTTAGGATCGACACC 1793916

RESULT 3
US-09-938-842A-4597
; Sequence 4597, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 68.9%; Score 18.6; DB 9; Length 2000;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 783 ACCCGCGCTTTAGGATCGACACTG 807

RESULT 4
US-09-938-842A-4597
; Sequence 4597, Application US/09938842A
; Patent No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 9; Length 2000;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 783 ACCCGCGCTTTAGGATCGACACTG 807

; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4597
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4597

Query Match
Best Local Similarity 66.7%; Score 18; DB 9; Length 1287;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACTG 26
   ||||| ||||| ||||| |||||
Db 90 CAGCGCGCTTTAGGATCGACACTG 65

RESULT 6
US-10-627-476-621/c
; Sequence 621, Application US/10627476
; Patent No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; FILE REFERENCE: BGI-125CPCN
```



```

; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1998-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 619
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1387)
; OTHER INFORMATION: RXN02233
; US-10-627-476-619

Query Match 66.7%; Score 18; DB 13; Length 1410;
Best Local Similarity 80.8%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
   ||| ||| ||| ||| ||| ||| |||
Db 190 CAGCGCGCTCTTTAGGAGCGACAACTG 165

RESULT 8
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-1

Query Match 66.7%; Score 18; DB 9; Length 3309400;
Best Local Similarity 80.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTG 26
   ||| ||| ||| ||| ||| ||| |||
Db 2048517 CAGCGCGCTCTTTAGGAGCGACAACTG 2048542

RESULT 9
US-10-138-970A-1
; Sequence 1, Application US/10138970A
; Publication No. US20030079255A1

```



```
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      65.2%; Score 17.6; DB 15; Length 9025608;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CCGCGCGCTTTAGGATCGACACCTG 26
    ||||| ||||| ||||| ||||| |||||
Db 8803965 CCGCGCTTCAGATCGTCACCGG 8803988

RESULT 13
US-10-027-632-324925/c
; Sequence 324925, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324925
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324925

Query Match      64.4%; Score 17.4; DB 13; Length 842;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
    ||||| ||||| ||||| ||||| |||||
Db 273 CACCCGCGCTTCAGGAACAACACGCTGA 247

RESULT 14
US-10-027-632-325019/c
; Sequence 325019, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 325019
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-325019

Query Match      64.4%; Score 17.4; DB 13; Length 842;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
    ||||| ||||| ||||| ||||| |||||
Db 273 CACCCGCGCTTCAGGAACAACACGCTGA 247

RESULT 15
US-10-027-632-324925/c
; Sequence 324925, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324925
; LENGTH: 842
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324925

Query Match      64.4%; Score 17.4; DB 16; Length 842;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CACCCGCGCTTTAGGATCGACACCTGA 27
    ||||| ||||| ||||| ||||| |||||
Db 273 CACCCGCGCTTCAGGAACAACACGCTGA 247

RESULT 16
US-10-027-632-325019/c
```

```

RESULT 18
US-10-424-599-103399/c
; Sequence 103399, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 103399
; LENGTH: 1063
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64388C.1
US-10-424-599-103399

Query Match          64.4%; Score 17.4; DB 13; Length 1063;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  CACCCGCGCTTTAGGATCGACACCTGA 27
          |||||
Db       447 CACCCGCGGTTCTGGATTGACGCGTGA 421
          |||||

RESULT 19
US-10-156-761-6630/c
; Sequence 6630, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6630
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1296)
US-10-156-761-6630

Query Match          64.4%; Score 17.4; DB 15; Length 1296;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  CACCCGCGCTTTAGGATCGACACCTGA 27
          |||||
Db       860 CAGCCGAGCTTGAGGTTGACACCCGGA 834
          |||||

RESULT 20
US-10-282-122A-26417/c

```

Sequence 26417, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA.034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 26417
 LENGTH: 1789
 TYPE: DNA
 ORGANISM: Mycobacterium bovis
 US-10-282-122A-26417

Query Match 64.4%; Score 17.4; DB 13; Length 1789;
 Best Local Similarity 77.8%; Pred. No. 1.7e+02;
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTAGGATCGACACCTGA 27
 DB 1293 CGCCCGCGCGCAGGATCGTCACCTTA 1267

RESULT 21
 US-09-712-363-135/C
 Sequence 135, Application US/09712363
 Patent No. US20020164589A1
 GENERAL INFORMATION:
 APPLICANT: Eisenberg, David
 APPLICANT: Rotstein, Sergio H.
 APPLICANT: Marcotte, Edward M.
 TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
 FILE REFERENCE: 07419-032001
 CURRENT APPLICATION NUMBER: US/09/712,363
 CURRENT FILING DATE: 2000-11-13
 PRIOR APPLICATION NUMBER: PCT/US00/02246
 PRIOR FILING DATE: 2000-01-28
 PRIOR APPLICATION NUMBER: 60/179,531

PRIOR FILING DATE: 2000-02-01
 PRIOR APPLICATION NUMBER: 60/117,844
 PRIOR FILING DATE: 1999-01-29
 PRIOR APPLICATION NUMBER: 60/118,206,
 PRIOR FILING DATE: 1999-02-01
 PRIOR APPLICATION NUMBER: 60/126,593
 PRIOR FILING DATE: 1999-03-26
 PRIOR APPLICATION NUMBER: 60/134,093
 PRIOR FILING DATE: 1999-05-14
 PRIOR APPLICATION NUMBER: 60/134,092
 PRIOR FILING DATE: 1999-05-14
 PRIOR APPLICATION NUMBER: 60/165,124
 PRIOR FILING DATE: 1999-11-12
 PRIOR APPLICATION NUMBER: 60/165,086
 PRIOR FILING DATE: 1999-11-12
 NUMBER OF SEQ ID NOS: 292
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 135
 LENGTH: 2283
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 US-09-712-363-135

Query Match 64.4%; Score 17.4; DB 9; Length 2283;
 Best Local Similarity 77.8%; Pred. No. 1.7e+02;
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTAGGATCGACACCTGA 27
 DB 1355 CGCCCGCGCGCAGGATCGTCACCTTA 1329

RESULT 22
 US-10-282-122A-28726/C
 Sequence 28726, Application US/10282122A
 Publication No. US20040029129A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Liangsu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari
 APPLICANT: Zyskind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.
 APPLICANT: Xu, H.
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA.034A
 CURRENT APPLICATION NUMBER: US/10/282,122A
 CURRENT FILING DATE: 2003-02-20
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/230,335
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: 60/230,347
 PRIOR FILING DATE: 2000-09-09
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/267,636
 PRIOR FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28726
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28726

Query Match 64.4%; Score 17.4; DB 13; Length 2283;

Best Local Similarity 77.8%; Pred. No. 1.7e+02;

Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCTGA 27

Db 1355 CGCCCGCGCGCAGGATCGTCACCTTA 1329

RESULT 23

US-09-974-300-1937/c

; Sequence 1937, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1937

; LENGTH: 999

; TYPE: DNA

; ORGANISM: Bacillus licheniformis

US-09-974-300-1937

Query Match

Best Local Similarity 63.7%; Score 17.2; DB 9; Length 999;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACA 22

Db 960 CACACGCGCTTTAGGATCCATA 939

RESULT 24

US-10-282-122A-23919

; Sequence 23919, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 23919

; LENGTH: 1257

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-10-282-122A-23919

Query Match

Best Local Similarity 63.7%; Score 17.2; DB 13; Length 1257;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CCGCGCTTTAGGATCGACACCT 25

Db 340 CCGCGCTTTGATCGACACCT 361

RESULT 25

US-10-369-493-45829

; Sequence 45829, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 45829

; LENGTH: 2493

; TYPE: DNA

; ORGANISM: Saccharomyces cerevisiae

US-10-369-493-45829

Query Match

Best Local Similarity 63.7%; Score 17.2; DB 16; Length 2493;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GCGCTTTAGGATCGACACCTGA 27

Db 1688 GCGCTTTAGGATCGCGCCCTGA 1709

RESULT 26

US-09-864-761-30891/c

; Sequence 30891, Application US/09864761

```
Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aescm.ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30891
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005669.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.44
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.5
; OTHER INFORMATION: SWISSPROT HIT: P23071, EVALUAE 2.40e-01
; OTHER INFORMATION: EST HUMAN HIT: AWS10940.1, EVALUAE 6.40e-02
; OTHER INFORMATION: NT HIT: AJ289880.1, EVALUAE 0.00e+00
US-09-864-761-30891
Query Match 63.0%; Score 17; DB 9; Length 279;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTG 26
Db 36 ACCCTCTCTTTGGGATCGACCACTG 12

RESULT 27
US-10-424-599-131520/c
; Sequence 131520, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 295684
; SEQ ID NO 131520
; LENGTH: 323
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_89770C.1
US-10-424-599-131520
Query Match 63.0%; Score 17; DB 13; Length 323;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCCGCGCTTTAGGATCGACACCTG 26
Db 131 ACCCGCGCTTTAGGATCGACACCTG 107

RESULT 28
US-09-728-445-661
; Sequence 661, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedzich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,358
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 661
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(349)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-661
Query Match 63.0%; Score 17; DB 9; Length 349;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25
Db 158 CACCCGCGCTCTCAGATCGACATCT 182

RESULT 29
US-10-424-599-108794
; Sequence 108794, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
```

APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
NUMBER FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 108794
LENGTH: 400
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_69254C.1
US-10-424-599-108794

Query Match 63.0%; Score 17; DB 13; Length 400;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CACCCGCGCTTTAGGATCGACACCT 25
||| ||||| ||||| ||||| |||||
Db 19 CAACTGTGTTTGGATCGAAACCT 43

RESULT 30
US-09-864-761-14328/c
Sequence 14328, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/532,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 14328
LENGTH: 463
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005669.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.53
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.44
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.5
US-09-864-761-14328

Query Match 63.0%; Score 17; DB 9; Length 463;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACCGGCGCTTTAGGATCGACACCTG 26
||| ||||| ||||| ||||| |||||
Db 262 ACCCTCTCTTTGGATCCCAACTG 238

Search completed: June 20, 2004, 17:45:28
Job time : 234.714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 543.069 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-13
Perfect score: 20
Sequence: 1 ggcgcgcacactcgactt 20

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_scs.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_scs.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rct.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_atgo_hum.*
- 40: em_atgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|------------|---------------------|
| 1 | 20 | 100.0 | 335 | 1 | MBO298910 | AJ298910 Mycobacte |
| 2 | 20 | 100.0 | 335 | 1 | MBO298911 | AJ298911 Mycobacte |
| 3 | 20 | 100.0 | 390 | 1 | MMI298907 | AJ298907 Mycobacte |
| 4 | 20 | 100.0 | 390 | 1 | MMI298909 | AJ298909 Mycobacte |
| 5 | 20 | 100.0 | 414 | 1 | MBO298905 | AJ298905 Mycobacte |
| 6 | 20 | 100.0 | 469 | 1 | MTU298906 | AJ298906 Mycobacte |
| 7 | 20 | 100.0 | 493 | 1 | MTU298908 | AJ298908 Mycobacte |
| 8 | 20 | 100.0 | 10894 | 1 | AE007121 | AE007121 Mycobacte |
| 9 | 20 | 100.0 | 318050 | 1 | EX248344 | EX248344 Mycobacte |
| 10 | 20 | 100.0 | 348676 | 15 | EX842581 | EX842581 Mycobacte |
| C 11 | 18.4 | 92.0 | 38477 | 1 | MSGB1723CS | L78825 Mycobacteri |
| C 12 | 18.4 | 92.0 | 42926 | 1 | MLCB1243 | AL023635 Mycobacte |
| C 13 | 18.4 | 92.0 | 348950 | 1 | MLEPRN7 | AL583923 Mycobacte |
| C 14 | 17 | 85.0 | 19632 | 1 | SNO224512 | AJ224512 Streptomy |
| C 15 | 17 | 85.0 | 67893 | 8 | AP003754 | AP003754 Oryza sat |
| C 16 | 17 | 85.0 | 137312 | 2 | AC135598 | AC135598 Oryza sat |
| C 17 | 17 | 85.0 | 158261 | 8 | CNS08C9S | AL772414 Oryza sat |
| C 18 | 17 | 85.0 | 308015 | 1 | AE016783 | AE016783 Pseudomon |
| C 19 | 16.8 | 84.0 | 276 | 8 | AF087589 | AF087589 Saxifraga |
| C 20 | 16.8 | 84.0 | 420 | 1 | CXSTPUC1 | X87993 C.xyli DNA |
| C 21 | 16.8 | 84.0 | 858 | 6 | AX122356 | AX122356 Sequence |
| C 22 | 16.8 | 84.0 | 858 | 6 | BD164473 | BD164473 Novel pol |
| C 23 | 16.8 | 84.0 | 968 | 8 | AK108178 | AK108178 Oryza sat |
| C 24 | 16.8 | 84.0 | 1654 | 3 | AF038597 | AF038597 Drosophill |
| C 25 | 16.8 | 84.0 | 1733 | 3 | AY069220 | AY069220 Drosophill |
| C 26 | 16.8 | 84.0 | 1755 | 3 | AY071071 | AY071071 Drosophill |
| C 27 | 16.8 | 84.0 | 2163 | 3 | DMU295829 | U29529 Drosophila |
| C 28 | 16.8 | 84.0 | 11366 | 1 | AE011952 | AE011952 Xanthomon |
| C 29 | 16.8 | 84.0 | 64195 | 2 | AP004394 | AP004394 Oryza sat |
| C 30 | 16.8 | 84.0 | 98286 | 2 | AP004394 | Continuation (33 o |
| C 31 | 16.8 | 84.0 | 123296 | 9 | HSB299818 | AL049694 Human DNA |
| C 32 | 16.8 | 84.0 | 148432 | 3 | AC004642 | AC004642 Drosophill |
| C 33 | 16.8 | 84.0 | 154840 | 3 | AC099018 | AC099018 Drosophill |
| C 34 | 16.8 | 84.0 | 157851 | 2 | AC020509 | AC020509 Drosophill |
| C 35 | 16.8 | 84.0 | 159478 | 2 | AL359638 | AL359638 Homo sapi |
| C 36 | 16.8 | 84.0 | 173613 | 3 | AC007475 | AC007475 Drosophill |
| C 37 | 16.8 | 84.0 | 184621 | 3 | AC007440 | AC007440 Drosophill |
| C 38 | 16.8 | 84.0 | 204697 | 2 | AC120995 | AC120995 Rattus no |
| C 39 | 16.8 | 84.0 | 222981 | 10 | AL669981 | AL669981 Mouse DNA |
| C 40 | 16.8 | 84.0 | 258635 | 2 | AC106208 | AC106208 Rattus no |
| C 41 | 16.8 | 84.0 | 292919 | 3 | AE003823 | AE003823 Drosophill |
| C 42 | 16.8 | 84.0 | 303823 | 3 | AE003462 | AE003462 Drosophill |
| C 43 | 16.8 | 84.0 | 337200 | 1 | AP005280 | AP005280 Corynebac |
| C 44 | 16.8 | 84.0 | 343243 | 1 | EX640414 | EX640414 Bordetell |
| C 45 | 16.8 | 84.0 | 347894 | 1 | EX640431 | EX640431 Bordetell |
| C 46 | 16.8 | 84.0 | 348706 | 1 | EX640445 | EX640445 Bordetell |
| C 47 | 16.8 | 84.0 | 349028 | 1 | EX640413 | EX640413 Bordetell |
| C 48 | 16.8 | 84.0 | 349737 | 1 | EX572597 | EX572597 Rhodopseu |
| C 49 | 16.8 | 84.0 | 349980 | 6 | AX127150 | AX127150 Sequence |
| C 50 | 16.4 | 82.0 | 1087 | 8 | AK107440 | AK107440 Oryza sat |
| C 51 | 16.4 | 82.0 | 1175 | 8 | AK107346 | AK107346 Oryza sat |
| C 52 | 16.4 | 82.0 | 1234 | 8 | AK108703 | AK108703 Oryza sat |
| C 53 | 16.4 | 82.0 | 1309 | 8 | AK055404 | AK055404 Oryza sat |
| C 54 | 16.4 | 82.0 | 1332 | 8 | OSHOX1 | X96881 Oryza sativ |
| C 55 | 16.4 | 82.0 | 1410 | 1 | AME277083 | AJ277083 Amycolato |
| C 56 | 16.4 | 82.0 | 1489 | 8 | AK105047 | AK105047 Oryza sat |
| C 57 | 16.4 | 82.0 | 1514 | 8 | AK064941 | AK064941 Oryza sat |
| C 58 | 16.4 | 82.0 | 1514 | 8 | AK100195 | AK100195 Oryza sat |
| C 59 | 16.4 | 82.0 | 1531 | 8 | AK122182 | AK122182 Oryza sat |
| C 60 | 16.4 | 82.0 | 1531 | 8 | AF211193 | AF211193 Oryza sat |
| C 61 | 16.4 | 82.0 | 3108 | 8 | EC013000 | EC013000 Homo sapi |
| C 62 | 16.4 | 82.0 | 3578 | 9 | AX676836 | AX676836 Sequence |
| C 63 | 16.4 | 82.0 | 3587 | 6 | AR366641 | AR366641 Sequence |
| C 64 | 16.4 | 82.0 | 3600 | 6 | BD141660 | BD141660 Use of Gl |
| C 65 | 16.4 | 82.0 | 3600 | 6 | BD141660 | BD141660 Use of Gl |

| | | | | | | | | | | | | | |
|-----|------|------|--------|---|--------------|---------------------|-----|------|------|--------|----|--------------|----------------------------------|
| 66 | 16.4 | 82.0 | 3600 | 6 | BD141661 | BD141661 Use of G1 | 139 | 15.8 | 79.0 | 1981 | 8 | AK107007 | AK107007 Oryza sat |
| 67 | 16.4 | 82.0 | 3600 | 6 | BD141662 | BD141662 Use of G1 | 140 | 15.8 | 79.0 | 2068 | 9 | BC033614 | BC033614 Homo sapi |
| 68 | 16.4 | 82.0 | 3600 | 6 | BD141663 | BD141663 Use of G1 | 141 | 15.8 | 79.0 | 2111 | 8 | AK100584 | AK100584 Oryza sat |
| 69 | 16.4 | 82.0 | 3600 | 6 | BD183558 | BD183558 Use of G1 | 142 | 15.8 | 79.0 | 2123 | 8 | AK071861 | AK071861 Oryza sat |
| 70 | 16.4 | 82.0 | 3600 | 6 | BD183559 | BD183559 Use of G1 | 143 | 15.8 | 79.0 | 2133 | 8 | BC004266 | BC004266 Homo sapi |
| 71 | 16.4 | 82.0 | 3600 | 6 | BD183660 | BD183660 Use of G1 | 144 | 15.8 | 79.0 | 2250 | 6 | AX748193 | AX748193 Sequence |
| 72 | 16.4 | 82.0 | 3600 | 6 | BD183661 | BD183661 Use of G1 | 145 | 15.8 | 79.0 | 2250 | 8 | AK093541 | AK093541 Homo sapi |
| 73 | 16.4 | 82.0 | 3600 | 6 | HSGLI | X07384 Human mRNA | 146 | 15.8 | 79.0 | 2250 | 8 | AK119966 | AK119966 Oryza sat |
| 74 | 16.4 | 82.0 | 4881 | 1 | AME318385 | AFJ318385 Amycolato | 147 | 15.8 | 79.0 | 2285 | 8 | AK072395 | AK072395 Homo sapi |
| 75 | 16.4 | 82.0 | 10163 | 9 | AFJ16573 | AFJ16573 Homo sapi | 148 | 15.8 | 79.0 | 2453 | 8 | AK074252 | AK074252 Homo sapi |
| 76 | 16.4 | 82.0 | 58349 | 8 | NCB5022 | AL355932 Neurospor | 149 | 15.8 | 79.0 | 2706 | 6 | AX713957 | AX713957 Sequence |
| 77 | 16.4 | 82.0 | 69187 | 8 | NCB9G16 | AL670003 Neurospor | 150 | 15.8 | 79.0 | 2706 | 6 | AX055912 | AX055912 Homo sapi |
| 78 | 16.4 | 82.0 | 76529 | 8 | NCB423H10 | AL42112 Oryza sat | 151 | 15.8 | 79.0 | 2859 | 6 | AX833932 | AX833932 Sequence |
| 79 | 16.4 | 82.0 | 97478 | 2 | AC105929 | AC105929 Magnapor | 152 | 15.8 | 79.0 | 2859 | 9 | AK096063 | AK096063 Homo sapi |
| 80 | 16.4 | 82.0 | 110000 | 2 | LMFLCHR15_5 | Continuation (6 of | 153 | 15.8 | 79.0 | 2894 | 9 | AX86063 | AX86063 Homo sapi |
| 81 | 16.4 | 82.0 | 110000 | 2 | LMFLCHR34_11 | Continuation (12 of | 154 | 15.8 | 79.0 | 4069 | 6 | AX882503 | AX882503 Sequence |
| 82 | 16.4 | 82.0 | 132927 | 8 | AC079890 | AC079890 Oryza sat | 155 | 15.8 | 79.0 | 4069 | 6 | BD159764 | BD159764 Primer fo |
| 83 | 16.4 | 82.0 | 136321 | 8 | AP003300 | AP003300 Oryza sat | 156 | 15.8 | 79.0 | 4069 | 6 | AK023003 | AK023003 Homo sapi |
| 84 | 16.4 | 82.0 | 150416 | 8 | OSJN00250 | AL731607 Oryza sat | 157 | 15.8 | 79.0 | 4318 | 9 | BC063786 | BC063786 Homo sapi |
| 85 | 16.4 | 82.0 | 155869 | 2 | AC103550 | AL303550 Oryza sat | 158 | 15.8 | 79.0 | 4611 | 9 | AB014564 | AB014564 Homo sapi |
| 86 | 16.4 | 82.0 | 157734 | 8 | OSJN00058 | AL606624 Oryza sat | 159 | 15.8 | 79.0 | 5241 | 9 | AK125717 | AK125717 Homo sapi |
| 87 | 16.4 | 82.0 | 164567 | 8 | OSJN00072 | AL606639 Oryza sat | 160 | 15.8 | 79.0 | 6773 | 3 | AY095005 | AY095005 Drosophil |
| 88 | 16.4 | 82.0 | 166987 | 2 | AC063917 | AC063917 Homo sapi | 161 | 15.8 | 79.0 | 7766 | 14 | AB096204 | AB096204 Cercopith |
| 89 | 16.4 | 82.0 | 169500 | 8 | AC115686 | AC115686 Oryza sat | 162 | 15.8 | 79.0 | 8144 | 6 | AX251388 | AX251388 Sequence |
| 90 | 16.4 | 82.0 | 172136 | 9 | AC022506 | AC022506 Homo sapi | 163 | 15.8 | 79.0 | 8913 | 6 | AX598871 | AX598871 Sequence |
| 91 | 16.4 | 82.0 | 185688 | 2 | AC018805 | AC018805 Homo sapi | 164 | 15.8 | 79.0 | 9305 | 1 | RRU310779 | RRU310779 Rhodospir |
| 92 | 16.4 | 82.0 | 192409 | 2 | AC072033 | AC072033 Homo sapi | 165 | 15.8 | 79.0 | 10519 | 3 | AF057167 | AF057167 Drosophil |
| 93 | 16.4 | 82.0 | 196921 | 2 | EX323601 | EX323601 Danio rer | 166 | 15.8 | 79.0 | 10519 | 6 | AX664476 | AX664476 Sequence |
| 94 | 16.4 | 82.0 | 197050 | 1 | AL646081 | AL646081 Ralstonia | 167 | 15.8 | 79.0 | 10786 | 1 | AE001871 | AE001871 Deinococc |
| 95 | 16.4 | 82.0 | 204050 | 1 | AL646073 | AL646073 Ralstonia | 168 | 15.8 | 79.0 | 11179 | 1 | AE004941 | AE004941 Pseudomon |
| 96 | 16.4 | 82.0 | 207782 | 8 | AP003764 | AP003764 Oryza sat | 169 | 15.8 | 79.0 | 11505 | 1 | AE011693 | AE011693 Xanthomon |
| 97 | 16.4 | 82.0 | 216050 | 1 | AL646076 | AL646076 Ralstonia | 170 | 15.8 | 79.0 | 11523 | 6 | AX251395 | AX251395 Sequence |
| 98 | 16.4 | 82.0 | 295150 | 1 | SC0939125 | SC0939125 Streptomy | 171 | 15.8 | 79.0 | 11523 | 6 | AX278012 | AX278012 Sequence |
| 99 | 16.4 | 82.0 | 298550 | 1 | AP005961 | AP005961 Bradyrhiz | 172 | 15.8 | 79.0 | 11523 | 6 | AX323709 | AX323709 Sequence |
| 100 | 16.4 | 82.0 | 300029 | 8 | AE017077 | AE017077 Oryza sat | 173 | 15.8 | 79.0 | 11523 | 6 | AX346824 | AX346824 Sequence |
| 101 | 16.4 | 82.0 | 300029 | 8 | AE017119 | AE017119 Oryza sat | 174 | 15.8 | 79.0 | 11523 | 6 | AX347440 | AX347440 Sequence |
| 102 | 16.4 | 82.0 | 300029 | 1 | AP005962 | AP005962 Bradyrhiz | 175 | 15.8 | 79.0 | 11523 | 6 | AX349161 | AX349161 Sequence |
| 103 | 16.4 | 82.0 | 300425 | 1 | AP005041 | AP005041 Streptomy | 176 | 15.8 | 79.0 | 11523 | 6 | AX657802 | AX657802 Sequence |
| 104 | 16.4 | 82.0 | 302998 | 1 | AE016921 | AE016921 Chromobac | 177 | 15.8 | 79.0 | 11523 | 6 | AX659076 | AX659076 Sequence |
| 105 | 16.4 | 82.0 | 344805 | 1 | EX840434 | EX840434 Bordetell | 178 | 15.8 | 79.0 | 11913 | 6 | AX795749 | AX795749 Sequence |
| 106 | 16.4 | 82.0 | 346879 | 1 | EX572598 | EX572598 Rhodosphe | 179 | 15.8 | 79.0 | 11913 | 6 | AX822257 | AX822257 Sequence |
| 107 | 16.4 | 82.0 | 348074 | 1 | EX640449 | EX640449 Bordetell | 180 | 15.8 | 79.0 | 11913 | 6 | AX825897 | AX825897 Sequence |
| 108 | 16.4 | 82.0 | 348134 | 1 | EX640420 | EX640420 Bordetell | 181 | 15.8 | 79.0 | 12262 | 1 | AE012405 | AE012405 Xanthomon |
| 109 | 16.4 | 82.0 | 349640 | 1 | EX572600 | EX572600 Rhodosphe | 182 | 15.8 | 79.0 | 12665 | 1 | AE012784 | AE012784 Chlorobiu |
| 110 | 16.4 | 82.0 | 349970 | 1 | EX571659 | EX571659 Wolinella | 183 | 15.8 | 79.0 | 19231 | 2 | AC132794 | AC132794 Continuation (7 of |
| 111 | 16.4 | 82.0 | 1378 | 8 | AK099104 | AK099104 Oryza sat | 184 | 15.8 | 79.0 | 22884 | 2 | AC090708 | AC090708 Homo sapi |
| 112 | 16.4 | 82.0 | 1472 | 8 | AK070401 | AK070401 Oryza sat | 185 | 15.8 | 79.0 | 25476 | 2 | LMFL236 | LMFL236 Homo sapi |
| 113 | 16.4 | 82.0 | 1617 | 6 | AX657577 | AX657577 Sequence | 186 | 15.8 | 79.0 | 70313 | 2 | AC016216 | AC016216 Homo sapi |
| 114 | 16.4 | 82.0 | 4965 | 1 | AB017350 | AB017350 Acidiphil | 187 | 15.8 | 79.0 | 81117 | 8 | AP005106 | AP005106 Oryza sat |
| 115 | 16.4 | 82.0 | 7171 | 1 | AF302797 | AF302797 Pseudomon | 188 | 15.8 | 79.0 | 82852 | 2 | AC017685 | AC017685 Drosophil |
| 116 | 16.4 | 82.0 | 12749 | 1 | AE012285 | AE012285 Xanthomon | 189 | 15.8 | 79.0 | 102784 | 2 | AC104708 | AC104708 Oryza sat |
| 117 | 16.4 | 82.0 | 37992 | 8 | EX842619 | EX842619 Neurospor | 190 | 15.8 | 79.0 | 107714 | 2 | AL356306 | AL356306 Homo sapi |
| 118 | 16.4 | 82.0 | 123570 | 8 | AP003804 | AP003804 Oryza sat | 191 | 15.8 | 79.0 | 110000 | 2 | AC132794 | AC132794 Continuation (6 of |
| 119 | 16.4 | 82.0 | 125454 | 8 | AP003374 | AP003374 Oryza sat | 192 | 15.8 | 79.0 | 110000 | 2 | LMFLCHR15_3 | LMFLCHR15_3 Continuation (4 of |
| 120 | 16.4 | 82.0 | 142549 | 2 | AP004271 | AP004271 Oryza sat | 193 | 15.8 | 79.0 | 110000 | 2 | LMFLCHR34_14 | LMFLCHR34_14 Continuation (15 of |
| 121 | 16.4 | 82.0 | 167203 | 8 | AP003264 | AP003264 Oryza sat | 194 | 15.8 | 79.0 | 119798 | 8 | CNS07YB6 | CNS07YB6 Oryza sat |
| 122 | 16.4 | 82.0 | 169032 | 2 | AC141741 | AC141741 Apis mell | 195 | 15.8 | 79.0 | 124266 | 8 | AC105744 | AC105744 Oryza sat |
| 123 | 16.4 | 82.0 | 207679 | 2 | EX511272 | EX511272 Danio rer | 196 | 15.8 | 79.0 | 124949 | 9 | AL358794 | AL358794 Human DNA |
| 124 | 16.4 | 82.0 | 273785 | 1 | SM5591793 | AL591793 Sinorhizo | 197 | 15.8 | 79.0 | 136263 | 8 | AP003201 | AP003201 Oryza sat |
| 125 | 16.4 | 82.0 | 299450 | 1 | AP005938 | AP005938 Bradyrhiz | 198 | 15.8 | 79.0 | 139434 | 2 | AP004754 | AP004754 Oryza sat |
| 126 | 16.4 | 82.0 | 347950 | 1 | AP003013 | AP003013 Mesorhizo | 199 | 15.8 | 79.0 | 141428 | 9 | AL139415 | AL139415 Human DNA |
| 127 | 16.4 | 82.0 | 348525 | 1 | EX640428 | EX640428 Bordetell | 200 | 15.8 | 79.0 | 143905 | 2 | AC137588 | AC137588 Oryza sat |
| 128 | 16.4 | 82.0 | 349028 | 1 | EX640413 | EX640413 Bordetell | 201 | 15.8 | 79.0 | 144141 | 8 | AP003196 | AP003196 Oryza sat |
| 129 | 16.4 | 82.0 | 349442 | 1 | EX640447 | EX640447 Bordetell | 202 | 15.8 | 79.0 | 147806 | 8 | AC119747 | AC119747 Genomic s |
| 130 | 15.8 | 79.0 | 247 | 6 | AX798739 | AX798739 Sequence | 203 | 15.8 | 79.0 | 151208 | 2 | AP003525 | AP003525 Oryza sat |
| 131 | 15.8 | 79.0 | 894 | 1 | AF146711 | AF146711 Chromobac | 204 | 15.8 | 79.0 | 151734 | 8 | OSJN00058 | OSJN00058 Oryza sat |
| 132 | 15.8 | 79.0 | 924 | 8 | AK111151 | AK111151 Oryza sat | 205 | 15.8 | 79.0 | 158805 | 2 | AC104847 | AC104847 Oryza sat |
| 133 | 15.8 | 79.0 | 1083 | 6 | AR386036 | AR386036 Sequence | 206 | 15.8 | 79.0 | 158911 | 8 | AC107619 | AC107619 Genomic s |
| 134 | 15.8 | 79.0 | 1207 | 1 | CHRPHEHYD | MS5915 Chromobacte | 207 | 15.8 | 79.0 | 159669 | 9 | AC104791 | AC104791 Homo sapi |
| 135 | 15.8 | 79.0 | 1268 | 8 | AK108031 | AK108031 Oryza sat | 208 | 15.8 | 79.0 | 159988 | 2 | AP005823 | AP005823 Oryza sat |
| 136 | 15.8 | 79.0 | 1595 | 8 | AK070686 | AK070686 Oryza sat | 209 | 15.8 | 79.0 | 161676 | 8 | AC135907 | AC135907 Genomic s |
| 137 | 15.8 | 79.0 | 1641 | 6 | BD131587 | BD131587 Method an | 210 | 15.8 | 79.0 | 161835 | 9 | AC098484 | AC098484 Homo sapi |
| 138 | 15.8 | 79.0 | 1744 | 8 | AK069652 | AK069652 Oryza sat | 211 | 15.8 | 79.0 | 169308 | 2 | AL591000 | AL591000 Homo sapi |

C 285 15.4 77.0 6972 9 HAAKTRSYV
 286 15.4 77.0 7011 6 AR198610
 287 15.4 77.0 7011 10 RATRBSI
 288 15.4 77.0 7420 9 HSKCNQ2
 289 15.4 77.0 8309 6 AR339592
 290 15.4 77.0 8902 2 AC017749
 291 15.4 77.0 8916 1 AC012206
 292 15.4 77.0 9695 10 AF055477
 293 15.4 77.0 10029 1 AF009640
 294 15.4 77.0 10155 1 AE014511
 295 15.4 77.0 10581 1 AE004701
 296 15.4 77.0 10982 1 AE002429
 297 15.4 77.0 11120 1 AE004734
 298 15.4 77.0 11219 1 AE001948
 299 15.4 77.0 11527 1 AE005953
 300 15.4 77.0 11819 1 AE011801

ACC09366 Drosophila
 BS000234 Pan trogl
 AC005696 Homo sapi
 AC128847 Rattus no
 AP003458 Oryza sat
 AP006440 Homo sapi
 AC092389 Oryza sat
 AL591107 Homo sapi
 AL124208 Rattus no
 AC099115 Rattus no
 AC119546 Rattus no
 AC117369 Rattus no
 AC127712 Rattus no
 AC096305 Rattus no
 AC133708 Rattus no
 AC130856 Rattus no
 AC107176 Rattus no
 AC133703 Rattus no
 AL391115 Streptomy
 AC095749 Rattus no
 AL591789 Sinorhizo
 AP005045 Streptomy
 AB017066 Oryza sat
 AB017115 Oryza sat
 AP006578 Gloeobact
 AP003597 Drosophila
 AL939111 Streptomy
 AC095174 Rattus no
 BX572605 Rhodospheu
 AR430774 Sequence
 AX028625 Sequence
 AX028589 Sequence
 AR430766 Sequence
 AX028609 Sequence
 AX056229 Sequence
 L27995 Brucella ab
 AX091219 Arabidops
 AX070030 Arabidops
 AJ330026 Homo sapi
 AX391540 Sequence
 AX391548 Sequence
 AK060334 Oryza sat
 AK104544 Oryza sat
 AR430774 Sequence
 AX028625 Sequence
 AF156692 Neisseria
 AR321631 Sequence
 AK063164 Oryza sat
 L79946 Homo sapien
 AF038173 Homo sapi
 AF052089 Homo sapi
 AF358663 Brucella
 AK101216 Oryza sat
 AK101162 Oryza sat
 M80792 Pseudomonas
 AK130880 Homo sapi
 AR430772 Sequence
 AX028621 Sequence
 AF277626 Synthetic
 AC014538 Drosophila
 AK120611 Oryza sat
 Y13070 S.coelicolo
 AK072069 Oryza sat
 AK067999 Oryza sat
 AB082951 Homo sapi
 X63453 D.melanogas
 BX649015 Homo sapi
 AB037724 Homo sapi
 BX537556 Homo sapi
 AX405958 Sequence
 X62666 Pseudomonas
 AY090863 Homo sapi
 AR304059 Sequence

ALIGNMENTS

RESULT 1
 MBO298910
 LOCUS MBO298910 335 bp DNA linear BCT 23-OCT-2001
 DEFINITION Mycobacterium bovis VNTR, strain 1503/96.
 ACCESSION AJ298910
 VERSION AJ298910.1 GI:16416365
 KEYWORDS repetitive element; tandem repeat; VNTR.
 SOURCE Mycobacterium bovis
 ORGANISM Mycobacterium bovis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.

REFERENCE 1
 AUTHORS Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
 TITLE Fluorescent VNTR (fVNTR) for improved high throughput molecular
 JOURNAL typing of Mycobacterium bovis
 REFERENCE 2 (bases 1 to 335)
 AUTHORS Llewellyn-Hughes, S.
 TITLE Direct Submission
 JOURNAL Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
 Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
 Surrey, KT15 3NB, UNITED KINGDOM

FEATURES
 source
 1. .335
 /organism="Mycobacterium bovis"
 /mol_type="genomic DNA"
 /strain="1503/96"
 /db_xref="taxon:1765"
 73..127
 /rpt_family="VNTR F2.1"
 /rpt_type=TANDEM
 128..206
 /rpt_family="VNTR F2.1"
 /rpt_type=TANDEM
 207..292
 /rpt_family="VNTR F2.1"
 /rpt_type=TANDEM
 repeat_region
 repeat_region
 repeat_region

ORIGIN
 Query Match 100.0%; Score 20; DB 1; Length 335;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCGCCGACACCTCGACTT 20
 |||||
 Db 37 GCGCCGACACCTCGACTT 56
 RESULT 2
 MBO298911
 LOCUS MBO298911 335 bp DNA linear BCT 23-OCT-2001

```

DEFINITION   Mycobacterium bovis VNTR, strain 4705.
ACCESSION    AJ298911.1 GI:16416366
VERSION      repetitive element; tandem repeat; VNTR.
KEYWORDS     Mycobacterium bovis
SOURCE       Mycobacterium bovis
ORGANISM     Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
              Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
AUTHORS      Unpublished
TITLE        Llewellyn-Hughes, S.
JOURNAL      Direct Submission
AUTHORS      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
JOURNAL      Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     Location/Qualifiers
             1..335
             /organism="Mycobacterium bovis"
             /mol_type="genomic DNA"
             /strain="4705"
             /db_xref="taxon:1765"
             repeat_region 73..127
             /rpt_family="VNTR F2.1"
             repeat_region 128..206
             /rpt_type="TANDEM"
             repeat_region 207..292
             /rpt_family="VNTR F2.1"
             /rpt_type="TANDEM"
ORIGIN
source
Query Match      100.0%; Score 20; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGCACCACTCGACTT 20
    |||||
Db 37 GCGCGCACCACTCGACTT 56

RESULT 3
MMI298907
LOCUS       Mycobacterium microti VNTR, strain 4777/97.
DEFINITION
ACCESSION  AJ298907
VERSION    AJ298907.1 GI:16416362
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium microti
ORGANISM   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
              Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
AUTHORS      Unpublished
TITLE        Llewellyn-Hughes, S.
JOURNAL      Direct Submission
AUTHORS      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
JOURNAL      Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     Location/Qualifiers
             1..390
             /organism="Mycobacterium microti"
             /mol_type="genomic DNA"
             /strain="4777/97"
             /db_xref="taxon:1806"
             repeat_region 73..127
             /rpt_family="VNTR F2.2"
             repeat_region 128..182
             /rpt_type="TANDEM"
             repeat_region 183..261
             /rpt_family="VNTR F2.2"
             /rpt_type="TANDEM"
ORIGIN
source
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGCACCACTCGACTT 20
    |||||
Db 37 GCGCGCACCACTCGACTT 56

RESULT 5
MMO298905
LOCUS       Mycobacterium microti VNTR, strain 4700/97.
DEFINITION
ACCESSION  AJ298909
VERSION    AJ298909.1 GI:16416364
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium microti
ORGANISM   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
              Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
AUTHORS      Unpublished
TITLE        Llewellyn-Hughes, S.
JOURNAL      Direct Submission
AUTHORS      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
JOURNAL      Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     Location/Qualifiers
             1..390
             /organism="Mycobacterium microti"
             /mol_type="genomic DNA"
             /strain="4700/97"
             /db_xref="taxon:1806"
             repeat_region 73..127
             /rpt_family="VNTR F2.2"
             repeat_region 128..182
             /rpt_type="TANDEM"
             repeat_region 183..261
             /rpt_family="VNTR F2.2"
             /rpt_type="TANDEM"
ORIGIN
source
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGCACCACTCGACTT 20
    |||||
Db 37 GCGCGCACCACTCGACTT 56

RESULT 5
MMO298905

```

```

DEFINITION   Mycobacterium bovis VNTR, strain 4705.
ACCESSION    AJ298911.1 GI:16416366
VERSION      repetitive element; tandem repeat; VNTR.
KEYWORDS     Mycobacterium bovis
SOURCE       Mycobacterium bovis
ORGANISM     Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
              Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
AUTHORS      Unpublished
TITLE        Llewellyn-Hughes, S.
JOURNAL      Direct Submission
AUTHORS      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
JOURNAL      Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     Location/Qualifiers
             1..335
             /organism="Mycobacterium bovis"
             /mol_type="genomic DNA"
             /strain="4705"
             /db_xref="taxon:1765"
             repeat_region 73..127
             /rpt_family="VNTR F2.1"
             repeat_region 128..206
             /rpt_type="TANDEM"
             repeat_region 207..292
             /rpt_family="VNTR F2.1"
             /rpt_type="TANDEM"
ORIGIN
source
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGCACCACTCGACTT 20
    |||||
Db 37 GCGCGCACCACTCGACTT 56

RESULT 4
MMI298909
LOCUS       Mycobacterium microti VNTR, strain 4700/97.
DEFINITION
ACCESSION  AJ298909
VERSION    AJ298909.1 GI:16416364
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium microti
ORGANISM   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
              Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
AUTHORS      Unpublished
TITLE        Llewellyn-Hughes, S.
JOURNAL      Direct Submission
AUTHORS      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
JOURNAL      Surrey, KT15 3NB, UNITED KINGDOM
FEATURES     Location/Qualifiers
             1..390
             /organism="Mycobacterium microti"
             /mol_type="genomic DNA"
             /strain="4700/97"
             /db_xref="taxon:1806"
             repeat_region 73..127
             /rpt_family="VNTR F2.2"
             repeat_region 128..182
             /rpt_type="TANDEM"
             repeat_region 183..261
             /rpt_family="VNTR F2.2"
             /rpt_type="TANDEM"
ORIGIN
source
Query Match      100.0%; Score 20; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGCGCACCACTCGACTT 20
    |||||
Db 37 GCGCGCACCACTCGACTT 56

RESULT 5
MMO298905

```

```

LOCUS          MBO298905          414 bp      DNA          linear          BCT 23-OCT-2001
DEFINITION     Mycobacterium bovis VNTR, strain AN 5.
ACCESSION      AJ298905
VERSION        AJ298905.1 GI:16416360
KEYWORDS       repetitive element; tandem repeat; VNTR.
SOURCE         Mycobacterium bovis
ORGANISM       Mycobacterium bovis
               Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
               Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
               tuberculosis complex.
REFERENCE      1 Llewellyn-Hughes,S., Llewellyn-Hughes,J. and Hewinson,R.G.
AUTHORS        Llewellyn-Hughes,S., Llewellyn-Hughes,J. and Hewinson,R.G.
TITLE          Fluorescent VNTR (fVNTR) for improved high throughput molecular
               typing of Mycobacterium bovis
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 414)
AUTHORS        Llewellyn-Hughes,S.
TITLE          Direct Submission
JOURNAL        Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
               Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
               Surrey, KT15 3NB, UNITED KINGDOM
FEATURES       Location/Qualifiers
               source
               1..414
               /organism="Mycobacterium bovis"
               /mol_type="genomic DNA"
               /strain="AN 5"
               /db_xref="taxon:1765"
               73..127
               /rpt_family="VNTR F2.1"
               /rpt_type=TANDEM
               repeat_region
               128..206
               /rpt_family="VNTR F2.1"
               /rpt_type=TANDEM
               repeat_region
               207..292
               /rpt_family="VNTR F2.1"
               /rpt_type=TANDEM
ORIGIN
Query Match          100.0%; Score 20; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GCGCGCACCACTCGACTT 20
        |||||
Db       37 GCGCGCACCACTCGACTT 56

RESULT 6
LOCUS          MTU298906          469 bp      DNA          linear          BCT 23-OCT-2001
DEFINITION     Mycobacterium tuberculosis VNTR, strain H37Rv.
ACCESSION      AJ298906
VERSION        AJ298906.1 GI:16416361
KEYWORDS       repetitive element; tandem repeat; VNTR.
SOURCE         Mycobacterium tuberculosis
ORGANISM       Mycobacterium tuberculosis
               Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
               Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
               tuberculosis complex.
REFERENCE      1 Llewellyn-Hughes,S., Llewellyn-Hughes,J. and Hewinson,R.G.
AUTHORS        Llewellyn-Hughes,S., Llewellyn-Hughes,J. and Hewinson,R.G.
TITLE          Fluorescent VNTR (fVNTR) for improved high throughput molecular
               typing of Mycobacterium bovis
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 469)
AUTHORS        Llewellyn-Hughes,S.
TITLE          Direct Submission
JOURNAL        Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
               Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
               Surrey, KT15 3NB, UNITED KINGDOM
FEATURES       Location/Qualifiers
               source
               1..469
               /organism="Mycobacterium tuberculosis"
               /mol_type="genomic DNA"
               /strain="H37Rv"
               /db_xref="taxon:1765"
               73..127
               /rpt_family="VNTR F4.1"
               /rpt_type=TANDEM
               repeat_region
               128..206
               /rpt_family="VNTR F4.1"
               /rpt_type=TANDEM
               repeat_region
               207..285
               /rpt_family="VNTR F4.1"
               /rpt_type=TANDEM
ORIGIN
Query Match          100.0%; Score 20; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GCGCGCACCACTCGACTT 20
        |||||
Db       37 GCGCGCACCACTCGACTT 56

RESULT 7
LOCUS          MBO298908          493 bp      DNA          linear          BCT 23-OCT-2001
DEFINITION     Mycobacterium bovis VNTR, strain 2259-93.
ACCESSION      AJ298908
VERSION        AJ298908.1 GI:16416363
KEYWORDS       repetitive element; tandem repeat; VNTR.
SOURCE         Mycobacterium bovis
ORGANISM       Mycobacterium bovis
               Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
               Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
               tuberculosis complex.
REFERENCE      1 Llewellyn-Hughes,S., Llewellyn-Hughes,J. and Hewinson,R.G.
AUTHORS        Llewellyn-Hughes,S., Llewellyn-Hughes,J. and Hewinson,R.G.
TITLE          Fluorescent VNTR (fVNTR) for improved high throughput molecular
               typing of Mycobacterium bovis
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 493)
AUTHORS        Llewellyn-Hughes,S.
TITLE          Direct Submission
JOURNAL        Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
               Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
               Surrey, KT15 3NB, UNITED KINGDOM
FEATURES       Location/Qualifiers
               source
               1..493
               /organism="Mycobacterium bovis"
               /mol_type="genomic DNA"
               /strain="2259.93"
               /db_xref="taxon:1765"
               73..127
               /rpt_family="VNTR F4.1"
               /rpt_type=TANDEM
               repeat_region
               128..206
               /rpt_family="VNTR F4.1"
               /rpt_type=TANDEM
               repeat_region
               207..285
               /rpt_family="VNTR F4.1"
               /rpt_type=TANDEM
ORIGIN
Query Match          100.0%; Score 20; DB 1; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GCGCGCACCACTCGACTT 20
        |||||
Db       37 GCGCGCACCACTCGACTT 56

RESULT 8

```


Mycobacterium bovis subsp. bovis AF2122/97
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,
Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,
Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,
Wheeler, P.R., Parkhill, J., Barrrell, B.G., Cole, S.T., Gordon, S.V. and
Hewinson, G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 (Microbiology)
2 (bases 1 to 318050)
Garnier, T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw/Adlestree,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. P74 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
location/Qualifiers
1. 318050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="Genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
complement(99..875)
/locus_tag="Mb2883c"
complement(99..875)
/locus_tag="Mb2883c"
/EC_number="1.-.-.-"
/notes="Mb2883c", len: 258 aa. Equivalent to Rv2857c.
len: 258 aa. from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 258 aa overlap). Probable short-chain
dehydrogenase/reductase (EC 1.-.-.-), highly similar to
various dehydrogenases e.g. O88068|SCI35.33c PROBABLE
DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor
(260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%
identity in 253 aa overlap); Q91376|PA1649 from
Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE
(253 aa). FASTA scores: opt: 569, E(): 2.1e-28, (39.2%
identity in 255 aa overlap); Q9EX74|MLHA SDR-LIKE ENZYME
from Rhodococcus erythropolis (246 aa), FASTA scores: opt:
567, E(): 2.8e-28, (41.15% identity in 248 aa overlap);
etc. Also similar to many Mycobacterium tuberculosis
dehydrogenases e.g. FABG3|Rv2002|MT2058|MTCV39.16c
PUTATIVE OXIDOREDUCTASE (260 aa). FASTA score: (38.3%
identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN
DEHYDROGENASES/REDUCTASES (SDR) FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE SHORT-CHAIN TYPE
DEHYDROGENASE/REDUCTASE"
/protein_id="CAD96569.1"
/db_xref="GI:31619829"
/translation="MDUSCSRLGRVAVITCGGSGIGLAGRRNRMRAGTIVGVDDV
EAGGAADSLSGLPFTVDVDSVANGFLDGAETIGRIDIAFNAGISPPENLNI
TELAQWRYQDVNLKSVYLCCLRALHVMVLAKGISVNTASVFAVMGSAISQISYTA
KGVLMQVQFARQIRVNALCPGVNTPLIQELFANPERAARRVHVPLGRF
AEPEETAAAVAFGLSDASFIATSFVLVDGGISSAYVTP"
complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"
complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"
complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"


```

/EC_number="1.2.1.3"
/notes="Mb2883c, aldC, len: 455 aa. Equivalent to RV2858c,
len: 455 aa. from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 455 aa overlap). Probable aldC.
aldehyde dehydrogenase (EC 1.2.1.3), similar to many e.g.
O88069|SCI35.34c PUTATIVE ALDEHYDE DEHYDROGENASE from
Streptomyces coelicolor (483 aa), FASTA scores: opt: 1872,
E(): 6.4e-109, (64.5% identity in 448 aa overlap);
Q9FAB1|ALDH1BT-ALDH ALDEHYDE DEHYDROGENASE from Bacillus
thermoleovorans (497 aa), FASTA scores: opt: 1157, E():
2.1e-64, (44.3% identity in 458 aa overlap); O33458|CYMC
P-CUMIC ALDEHYDE DEHYDROGENASE from Pseudomonas putida
(494 aa), FASTA scores: opt: 1149, E(): 6.5e-64, (43.15%
identity in 452 aa overlap);
P40047|DHAS YEAST|ALDS|ALDH5|ALD3|YERO73W ALDEHYDE
DEHYDROGENASE from Saccharomyces cerevisiae (Baker's
yeast) (519 aa), FASTA scores: opt: 1091, E(): 2.7e-60,
(38.55% identity in 459 aa overlap);
P80668|FEAB ECOLI|PADA|MAOB|B1385 PHENYLACETALDEHYDE
DEHYDROGENASE (EC 1.2.1.39) from Escherichia coli strain
K12 (499 aa), FASTA scores: opt: 1074, E(): 3e-59, (42.2%
identity in 462 aa overlap); etc. Also similar to many M.
tuberculosis dehydrogenases e.g. P71823|RV0768|MTCY369.13
(489 aa), FASTA score: (38.1% identity in 467 aa overlap).
Contains P500687 Aldehyde dehydrogenases glutamic acid
active site and P500070 Aldehyde dehydrogenases cysteine
active site. BELONGS TO THE ALDEHYDE DEHYDROGENASES
FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE ALDEHYDE DEHYDROGENASE ALDC"
/protein_id="CAD96570.1"
/db_xref="GI:31619630"
/complement(2236, 3162)
/locus_tag="Mb2884c"
/complement(2235, 3162)
/locus_tag="Mb2884c"
/EC_number="2.3.5.-"
/EC_number="6.3.5.-"
/notes="Mb2884c, -, len: 308 aa. Equivalent to RV2859c,
len: 308 aa. from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 308 aa overlap). Possible
amidotransferase (EC 6.3.5.- or 2.-.-.-) equivalent (but
longer 58 aa) to Q9CBU9|ML1573 POSSIBLE AMIDOTRANSFERASE
from Mycobacterium leprae (249 aa), FASTA scores: opt:
1226, E(): 3e-64, (71.55% identity in 239 aa overlap).
Also similar to other amidotransferases and hypothetical
proteins, but shorter in N-terminus e.g. O88072|SCI35.37
HYPOTHETICAL 25.3 KDA PROTEIN from Streptomyces coelicolor
(242 aa), FASTA scores: opt: 683, E(): 1.2e-32, (47.65%
identity in 235 aa overlap); AAK79730|Q97188|CAC1764
PREDICTED GLUTAMINE AMIDOTRANSFERASE from Clostridium
acetobutylicum (241 aa), FASTA scores: opt: 458, E():
1.6e-19, (32.9% identity in 246 aa overlap);
AAK75201|Q97QV9|SPI089 GLUTAMINE AMIDOTRANSFERASE CLASS I
from Streptococcus pneumoniae (229 aa), FASTA scores: opt:
431, E(): 5.6e-18, (34.75% identity in 236 aa overlap);
etc. Contains three 17 aa repeats at the N-terminus very
similar to those in other Mycobacterium tuberculosis
proteins e.g. Q10699|YJ30_MYCTU|RV2090|MT2151|MTCY49.30
PUTATIVE 5'-3' EXONUCLEASE RV2090 (EC 3.1.11.-).
/codon_start=1
/transl_table=11
/product="POSSIBLE AMIDOTRANSFERASE"

```

```

/protein_id="CAD96571.1"
/db_xref="GI:31619631"
/transl_table=11
/translation="MDLSASRSDGGDPLRPASPRLSRSPVSDGDPDLPASPRLSRSPVS
DGGDPLRPASPRLSRSPVSGLTAYLEOVRTGWDIIPAGYLPADYFEGITVAGG
VAVLLPQVDPDESVCVLDLHALVITGGYDLDPAAAYQEPHDPATDPRPGRDAWEF
ALLRQALQGMPLVIGICRTQVNLVALGSLTHQLPLDILGHSGHAGNAGVFRPLPVHT
AGSTRALAEIGESADVPCHVHQIDQVGGGLVVSADVVDVGVIEALELPDGTFLAVQW
HFEKSDDURLFKALVDAAAGYAGRSQAEPR"
/complement(3143, 4516)
/gene="glnA4"
/locus_tag="Mb2885c"
/complement(3143, 4516)
/gene="glnA4"
/locus_tag="Mb2885c"
/EC_number="6.3.1.2"
/notes="Mb2885c, glnA4, len: 457 aa. Equivalent to RV2860c,
len: 457 aa. from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 457 aa overlap). Probable glnA4.
glutamine synthetase class II (EC 6.3.1.2), similar to
many glutamine synthetases e.g. O88070|SCI35.35c from
Streptomyces coelicolor (462 aa), FASTA scores: opt: 1947,
E(): 8.2e-120, (64.15% identity in 452 aa overlap);
Q98H15|ML13074 from Rhizobium loti (Mesorhizobium loti)
(465 aa), FASTA scores: opt: 1321, E(): 7.8e-79, (46.7%
identity in 452 aa overlap); Q98EM0|MLL4187 from Rhizobium
loti (Mesorhizobium loti) (456 aa), FASTA scores: opt:
699, E(): 4.6e-38, (33.5% identity in 454 aa overlap);
Q9CDL9|GLNA from Lactococcus lactis (subsp. lactis)
(Streptococcus lactis) (446 aa), FASTA scores: opt: 633,
E(): 8.2e-34, (32.45% identity in 456 aa overlap); etc.
Also similar to three other potential glutamine synthetases
in Mycobacterium tuberculosis:
Q10378|GLN2_MYCTU|GLN22|Rv2222c|MTCY190.33c|MTCY427
.03c PROBABLE GLUTAMINE SYNTHETASE (446 aa), FASTA score:
(31.1% identity in 453 aa overlap); RV1878|glnA3 and
RV2220|glnA1. BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE GLUTAMINE SYNTHETASE GLNA4 (GLUTAMINE
SYNTHASE) (GS-II)"
/protein_id="CAD96572.1"
/db_xref="GI:31619632"
/translation="MTGPGSPPLAWTELRVLAAGDVDTVI VAFDMDQRLAGKRISG
RHFVDIATRGVECCSYLLAVDLNTVGYAMASMDTGYDMVMTDLSLRLPLWL
PGTALVIADLVWADGEVAVSPRSILRDLRLKARGLVADVATELETFVFPQYRQA
WASGEVGLT'PASDNYIDYAILASSMEPLRLDIRLGMAGAGLRFEAVGECNMGQOEI
GRFYDALVTCNHNHAIYKNGAKEIADQHGKSLTFYAKYDEREGNSCHLHVSIRGTDGS
AVFADNSNGPHGSSMFRSFVAGQLATREFTLCYAFPTINSYKRFADSSFAPTALAWGL"
Query Match 100.0%; Score 20; DB 1; Length 318050;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCCTCCGACACCTCGACTT 20
Db 71063 GCCTCCGACACCTCGACTT 71082
RESULT 10
ID BX842581 standard; circular genomic DNA; PRO; 348676 BP.
XX BX842581
AC BX842581; AL008883; AL008967; AL021070; AL021287; AL021309; AL123456;
AC Z74024; Z74697; Z81331; Z83018; Z83857; Z83858; Z83866; Z95150; Z95207;
XX BX842581.1
SV BX842581.1
DT 21-NOV-2003 (Rel. 77, Created)
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)
DE Mycobacterium tuberculosis H37Rv complete genome; segment 10/13
XX complete genome.

```


FT repeat_unit
FT complement(2845. .2860)
FT /note="16 bp partial direct repeat, GTGCTGAGACCAAAA, of
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(2901. .2936)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(2976. .3011)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3049. .3084)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3121. .3156)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3193. .3228)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3266. .3301)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3338. .3373)
FT /note="36 bp direct repeat, 32 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3399. .3434)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3476. .3511)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3548. .3583)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3624. .3659)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3701. .3736)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGAGACCAAAAACCCGAGAGGGGACGGAAC"
FT complement(3774. .3809)

Query Match 100.0%; Score 20; DB 15; Length 348676;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGGACCACTCGACTT 20
DB 120398 GCGCGGACCACTCGACTT 120417

RESULT 11
MSGB1723CS/c 38477 bp DNA linear BCT 17-DEC-2001
LOCUS Mycobacterium leprae cosmid B1723 DNA sequence.
DEFINITION L78825
ACCESSION L78825
VERSION L78825.1 GI:1377772
KEYWORDS
SOURCE Mycobacterium leprae
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE 1 (sites)
AUTHORS Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
TITLE Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae
JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)
MEDLINE 93188700
PUBMED 8446027
REFERENCE 2 (bases 1 to 38477)
AUTHORS Smith, D.R., Richterich, P., Rubenfield, M., Rice, P.W., Butler, C., Lee, H.M., Kirst, S., Gundersen, K., Abendschan, K., Xu, Q., Chung, M., Deloughery, C., Aldredge, T., Maher, J., Lundstrom, R., Tulig, C.,

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Falls, K., Imrich, J., Torrey, D., Engelstein, M., Breton, G., Madan, D., Nietupski, R., Seitz, B., Mao, J. I. et al.
Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome
Genome Res. 7 (8), 802-819 (1997)
97413161
9267804

Original source text: Mycobacterium leprae (clone: cosmid B1723)
(tissue library: Loris 6) DNA.

This sequence was generated by the Genome Sequencing Center at
Genome Therapeutics Corporation (Collaborative Research Division),
100

Reaver St., Waltham, MA, 02154. Please contact Doug Smith
(smith@eric.com) for further information. The sequence represents
the

insert of a Loris 6 cosmid clone from a mapped set of clones
constructed from M. leprae genomic DNA isolated from armadillo
liver

[3]. The sequence may not represent the entire cloned insert of
the

cosmid if an overlapping region was previously sequenced from
another

clone. Coding sequences larger than 60 amino acids were predicted
on

the basis of codon usage and homology information. An attempt was
made

to locate the most probable start site based on codon usage,
homology,

the presence of a Shine-Dalgarno sequence, or overlapping orf that
suggested translational coupling. It is possible that the actual
start

site differs from the one selected.

FEATURES
source

Location/Qualifiers

1..38477

/organism="Mycobacterium leprae"

/mol_type="genomic DNA"

/specific_host="Dasyptus novemcinctus"

/db_xref="taxon:1769"

/clone="cosmid B1723"

/tissue_lib="Loris 6"

/note="The liver of the armadillo was used to isolate the
Mycobacterium leprae."

ORIGIN

Query Match 92.0%; Score 18.4; DB 1; Length 38477;
Best Local Similarity 95.0%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGGACCACTCGACTT 20
DB 23828 GCGCGGACCACTCGACTT 23809

RESULT 12

MLCB1243/c

LOCUS MLCB1243

DEFINITION Mycobacterium leprae cosmid B1243.

ACCESSION AL023635

VERSION AL023635.1

KEYWORDS GI:3150214

accd6; aceB; Acetyl/Propionyl CoA Carboxylase; acpM; acyl carrier

protein; ATP-dependent DNA helicase; beta-ketoacyl-ACP synthase;

formamidopyrimidine-dna glycosylase; fpg; kasA; kasB; kdsB;

lipopolysaccharide core biosynthesis protein; malonyl CoA-acyl

carrier protein transacylase; mutM; oxidoreductase; pyruvate

dehydrogenase ei component; recG; ribonuclease III; RUF; rnc.

Mycobacterium leprae

Mycobacterium leprae

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

1 (bases 1 to 42926)

Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.

Use of an ordered cosmid library to deduce the genomic organization

of Mycobacterium leprae

Mol. Microbiol. 7 (2), 197-206 (1993)

JOURNAL

| | | | | |
|-------------|--|--|--|--|
| CDS | /gene="MLCB1243.03c" /note="PS00062 Aldo/keto reductase family signature 2" complement(3448..4254) /note="MLCB1243.04c, probable oxidoreductase pseudogene, similar to e.g. 2PKG CORSP 2,5-diketo-d-gluconic acid reductase and to downstream ORF MLCB1243.03c" /pseudo /codon_start=1 /transl_table=11 /complement(3450..3453) /note="possible RBS upstream of MLCB1243.03c" 4318..5383 | | JOURNAL COMMENT | |
| | /note="MLCB1243.05, probable pseudogene, equivalent to M. tuberculosis hypothetical lipase/esterase TR:P95125 (EMBL:Z83018) MTCV349.17 (376 aa)" /pseudo /codon_start=1 /transl_table=11 5457..5460 /note="possible RBS upstream of MLCB1243.06" 5472..6248 /gene="MLCB1243.06" 5472..6248 /gene="MLCB1243.06" /note="MLCB1243.06, possible transmembrane protein, len: 258 aa; equivalent to M. tuberculosis hypothetical membrane protein TR:Q33272 (EMBL:Z83018) MTCV349.18 (255 aa), fasta scores; Opt: 1097 z-score: 1595.4 E(): 0, 65.5% identity in258 aa overlap" /codon_start=1 /transl_table=11 /product="putative transmembrane protein" /protein_id="CAA19187.1" /db_xref="GI:3150218" /db_xref="GOA:O69463" | | Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk Notes: Details of M. leprae sequencing at the Sanger Centre are available from http://www.sanger.ac.uk/Projects/M_leprae/ A relational database containing the M. leprae sequences is available from http://genolist.pasteur.fr/Leprae/ . | |
| RBS | /note="possible RBS upstream of MLCB1243.06" 5472..6248 /gene="MLCB1243.06" 5472..6248 /gene="MLCB1243.06" /note="MLCB1243.06, possible transmembrane protein, len: 258 aa; equivalent to M. tuberculosis hypothetical membrane protein TR:Q33272 (EMBL:Z83018) MTCV349.18 (255 aa), fasta scores; Opt: 1097 z-score: 1595.4 E(): 0, 65.5% identity in258 aa overlap" /codon_start=1 /transl_table=11 /product="putative transmembrane protein" /protein_id="CAA19187.1" /db_xref="GI:3150218" /db_xref="GOA:O69463" | | FEATURES source repeat_region gene CDS | |
| CDS | /note="possible RBS upstream of MLCB1243.06" 5472..6248 /gene="MLCB1243.06" 5472..6248 /gene="MLCB1243.06" /note="MLCB1243.06, possible transmembrane protein, len: 258 aa; equivalent to M. tuberculosis hypothetical membrane protein TR:Q33272 (EMBL:Z83018) MTCV349.18 (255 aa), fasta scores; Opt: 1097 z-score: 1595.4 E(): 0, 65.5% identity in258 aa overlap" /codon_start=1 /transl_table=11 /product="putative transmembrane protein" /protein_id="CAA19187.1" /db_xref="GI:3150218" /db_xref="GOA:O69463" | | misc_feature | |
| Query Match | 92.0%; Score 18.4; DB 1; Length 42926; Best Local Similarity 95.0%; Pred. No. 5.3e+02; Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | Gene | |
| Db | 1 GCGCGCACCACTCGACTT 20 15698 GCGCGCACCACTCTACTT 15679 | | CDS | |
| RESULT 13 | MLEPRTN7 348950 bp DNA linear BCT 20-FEB-2001 | | Gene | |
| LOCUS | Mycobacterium leprae strain TN complete genome; segment 7/10. | | CDS | |
| DEFINITION | AL583923 AL450380 | | Gene | |
| ACCESSION | AL583923.1 GI:13093428 | | CDS | |
| VERSION | AL583923.1 GI:13093428 | | Gene | |
| KEYWORDS | Mycobacterium leprae | | CDS | |
| SOURCE | Mycobacterium leprae | | Gene | |
| ORGANISM | Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium. | | CDS | |
| REFERENCE | 1 (bases 1 to 348950) | | Gene | |
| AUTHORS | Cole, S.T., Eigmler, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Felgwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jags, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R. and Barrall, B.G. | | Gene | |
| TITLE | Massive gene decay in the leprosy bacillus | | CDS | |
| JOURNAL | Nature 409 (6823), 1007-1011 (2001) | | Gene | |
| MEDLINE | 21128732 | | CDS | |
| PUBMED | 11234002 | | Gene | |
| REFERENCE | 2 (bases 1 to 348950) | | CDS | |
| AUTHORS | Parkhill, J. | | Gene | |
| TITLE | Direct Submission | | CDS | |